LITHUANIAN UNIVERSITY OF HEALTH SCIENCES

Greta Varkalaitė

COMPREHENSIVE ANALYSIS OF CIRCULATING NUCLEIC ACIDS IN GASTRIC CANCER PATIENTS' BLOOD PLASMA

Doctoral Dissertation
Natural Sciences,
Biology (N 010)

Kaunas, 2022

Dissertation has been prepared at the Laboratory of Clinical and Molecular Gastroenterology of Institute for Digestive Research of Lithuanian University of Health Sciences during the period of 2017–2021.

Scientific Supervisor:

Prof. Dr. Jurgita Skiecevičienė (Lithuanian University of Health Sciences, Natural Sciences, Biology – N 010).

Dissertation is defended at the Biology Research Council of the Lithuanian University of Health Sciences:

Chairperson

Prof. Dr. Edgaras Stankevičius (Lithuanian University of Health Sciences, Natural Sciences, Biology – N 010);

Members:

Dr. Povilas Ignatavičius (Lithuanian University of Health Sciences, Medical and Health Sciences, Medicine – M 001);

Prof. Dr. Rasa Banienė (Lithuanian University of Health Sciences, Natural Sciences, Biology – N 010);

Prof. Dr. Arvydas Lubys (Vilnius University, Natural Sciences, Biology – N 010);

Prof. Dr. Ants Kurg (University of Tartu (Estonia), Natural Sciences, Biology – N 010).

Dissertation will be defended at the open session of the Biology Research Council of the Lithuanian University of Health Sciences on the 4th of July, 2022 at 2 p.m. in the A-202 auditorium of "Santaka" Valley Centre for the Advanced Pharmaceutical and Health Technologies of Lithuanian University of Health Sciences.
Address: Sukilėlių pr. 13, LT-50162 Kaunas, Lithuania.

LIETUVOS SVEIKATOS MOKSLŲ UNIVERSITETAS

Greta Varkalaitė

IŠSAMI LAISVAI CIRKULIUOJANČIŲ NUKLEORŪGŠČIŲ ANALIZĖ SKRANDŽIO VĖŽIU SERGANČIŲ PACIENTŲ KRAUJO PLAZMOJE

Daktaro disertacija
Gamtos mokslai,
biologija (N 010)

Kaunas, 2022

Disertacija rengta 2017–2021 metais Lietuvos sveikatos mokslų universiteto Virškinimo sistemos tyrimų instituto Klinikinės ir molekulinės gastroenterologijos laboratorijoje.

Mokslinis vadovas

Prof. Dr. Jurgita Skiecevičienė (Lietuvos sveikatos mokslų universitetas, gamtos mokslai, biologija – N 010).

Disertacija ginama Lietuvos sveikatos mokslų universiteto biologijos mokslo krypties taryboje:

Pirmininkas

prof. dr. Edgaras Stankevičius (Lietuvos sveikatos mokslų universitetas, gamtos mokslai, biologija – N 010);

Nariai:

dr. Povilas Ignatavičius (Lietuvos sveikatos mokslų universitetas, medicinos ir sveikatos mokslai, medicina – M 001);

prof. dr. Rasa Banienė (Lietuvos sveikatos mokslų universitetas, gamtos mokslai, biologija – N 010);

prof. dr. Arvydas Lubys (Vilniaus Universitetas, gamtos mokslai, biologija – N 010);

prof. dr. Ants Kurg (Tartu Universitetas (Estija), gamtos mokslai, biologija – N 010).

Disertacija bus ginama viešajame Biologijos mokslo krypties tarybos posėdyje 2022 m. liepos 4 d. 14 val. Lietuvos sveikatos mokslų universiteto „Santakos" slėnio Naujausių farmacijos ir sveikatos technologijų centro A-202 auditorijoje.

Disertacijos gynimo vietos adresas: Sukilėlių pr. 13, LT-50162 Kaunas.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

(V)AF – (variant) allele frequency
AG – atrophic gastritis
AUC – area under the curve
CAV1 – Caveolin 1
CDH1 – Cadherin 1
cfDNA – cell-free DNA
cfNA – cell-free nucleic acids
CON – control
COSMIC – Catalogue of Somatic Mutations in Cancer
ctDNA – circulating tumour DNA
DOT1-L – DOT1 Like Histone Lysine Methyltransferase
ELISA – enzyme-linked immunosorbent assay
FBXO24 – F-Box Protein 24
FITC – Fluorescein isothiocyanate
GAPDH – Glyceraldehyde-3-Phosphate Dehydrogenase
GC – gastric cancer
IRF1 – Interferon Regulatory Factor 1
MDS – multidimensional scaling analysis
miRNA – microRNA
mRNA – messenger RNA
MSI – microsatellite instability
NGS – next-generation sequencing
OLGA – Operative Link on Gastritis Assessment
PALB2 – Partner and Localizer of BRCA2
PTEN – Phosphatase and tensin homolog
ROC – receiver operating characteristic
smRNA – small RNA
TMB – tumour mutational burden
TP53 – tumour protein p53
TSC1 – TSC Complex Subunit 1
TXNIP – Thioredoxin Interacting Protein
UMI – unique molecular identifier
UTR – untranslated region
WB – Western Blot
WBC – white blood cell
WES – whole exome sequencing
WGS – whole genome sequencing

INTRODUCTION

It was shown in year 1948 by Mandel and Métais that fragments of cell-free nucleic acids (cfNAs) could be detected in the bloodstream [1]. Later, in year 1991 the term "liquid biopsy" was introduced, when Sidransky *et al*. published results that DNA from urinary sediments of patients with invasive bladder cancer carried mutations in *TP53* gene [2]. In the following decades, improving technologies led to a great variety of studies analysing possible cfNA application: from non-invasive prenatal testing (NIPT) and organ transplantation monitoring [3–5] to detection and monitoring of many cancers [6]. Today, liquid biopsy is understood as an analysis of blood and other body fluids obtained in a non-invasive or minimally invasive manner to test the molecular landscape of solid tumours [7].

cfNAs are released during the cell apoptosis, necrosis, or even active secretion, and reflects the genetics and epigenetics of the cell of origin [8–11] (Fig. 1.1.) Studies involving healthy individuals show that the highest fraction is composed of cfDNA originating from haematopoietic cells [10], and cfDNA yield could increase after intense exercise, traumas, or tissue injury. Typical cfDNA size showed by gel electrophoresis is considered to be around 180 bp [12–14], and corresponds to the length of DNA wrapped around a nucleosome plus linker DNA [15]. However, different fragmentation could be also observed: shorter fragments were detected in case of a tumour, fetal derived cfDNA [13,16], and organ transplantation [17]. Specific fragmentation could be explained by differences in nucleosome wrapping and nuclease activity in different cells, especially cancer. After the cfDNA enters the blood circulation, the clearance is rapid: researches including analysis of Epstein-Barr virus and maternal DNA suggests that the half-life of cfDNA in the circulation is between 16 minutes and 2.5 hours [18–20]. It is important to note that this feature is considered as one of the most important characteristics of cfDNA especially for the "real-time snapshot" applications and longitudinal sampling in patients' care.

Regular monitoring of a disease state is extremely necessary for cancer patients' care. Thus, tumour-derived cfDNA – named circulating tumour DNA (ctDNA) – could be a very promising molecular tool that would minimize the need of invasive interventions. ctDNA may originate from primary or metastatic lesions and circulating tumour cells. According to the literature the fraction of ctDNA varies from less than 0.1 % to more than 25 % [21,22]. Pioneer studies started with an evaluation of cfDNA and/or ctDNA yield dynamics. It was associated with tumour burden, reoccurrence, and response to therapy [21,23,24]. Recent technological improvement redirected a focus to

a deeper analysis of ctDNA. The development of next-generation sequencing (NGS) enabled not only single-gene and/or variant detection in the ctDNA but also the use of targeted panels or even whole-genome sequencing (WGS), and identification of a wide range of alterations including chromosomal aberrations, amplification, and gene rearrangements [6,25–28].
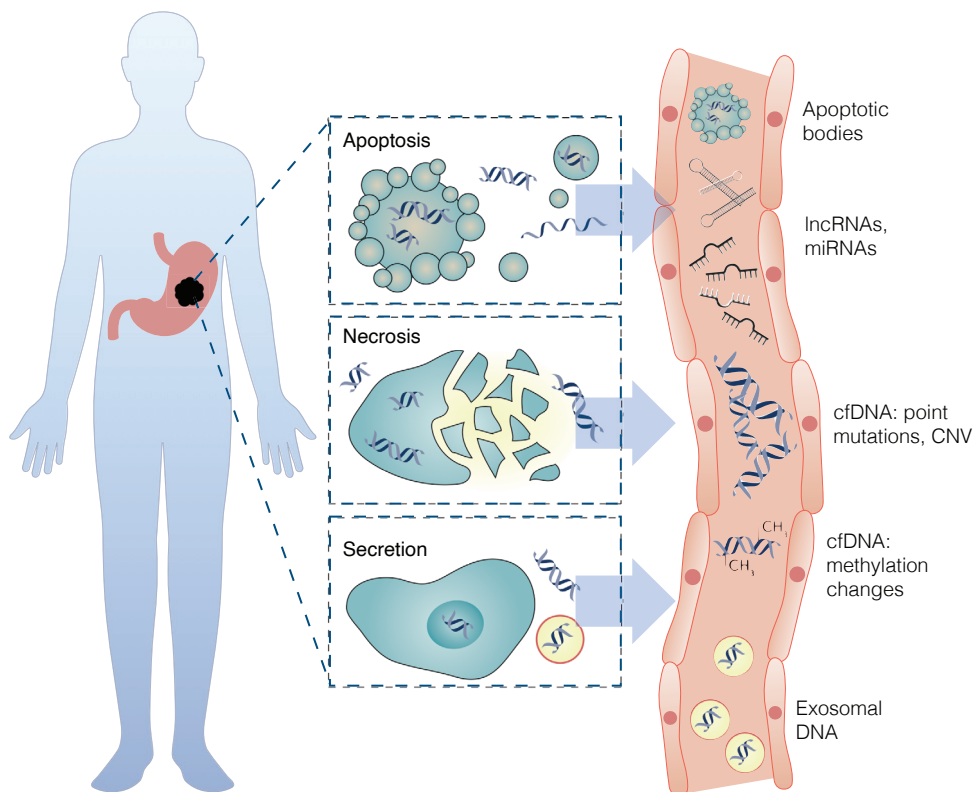

Fig. 1.1. *Origins and range of alterations in cell-free nucleic acids.*

Tumour cells release cfNA through a combination of apoptosis, necrosis, and secretion. In blood stream wide range of tumour-derived molecules and molecular aberrations could be detected. Adapted from Jonathan C. M. Wan, et al. 2017.

Alterations detected in circulating DNA represent genetic changes, which are only a subset of tumour-related molecular information. Besides the possibility to analyse the methylation status of cfDNA, blood-based microRNA (miRNA) profiling can also provide highly valuable epigenetic information. To date, more than 2 500 miRNAs are catalogued and associated with important biological processes (miRBase version 22, December 2021) [29]. A significant number of human transcripts are regulated by miRNAs showing the important role of these molecules in many physiological and pathological

processes. Moreover, several characteristics make miRNA a great biomarker candidate. First, miRNAs are one of the most abundant RNA molecules in the blood. Also, the small size of miRNA molecules enables its reliable analysis in many different biological sources such as blood (total blood, plasma, or serum), circulating exosomes, and even urine, saliva, and sputum [30]. Studies conducted by Lawrie et *al.* and Ghafouri-Fard *et al.* suggest that the profile of circulating miRNAs mirrors the profiles of tumour tissues, and could be implied for minimally invasive diagnostics [31,32]. However, not only diagnostic but also therapeutic potential of miRNAs is important, there are several 1 and 2 phase trials of miRNA-based therapy for hepatitis C, hereditary nephritis, and other [33]. For this reason, functional analysis involving target prediction, and *in vitro* assays are of great importance.

Gastric cancer (GC) is one of the most common oncological diseases of the gastrointestinal tract in Lithuania and around the globe [34] (Fig. 1.2.). It is frequently diagnosed at the advanced stages and resulted in 768 000 deaths globally in 2020 (2020 GLOBOCAN data), which makes GC the third most common cause of cancer-related deaths [34]. In most cases GC development is a stepwise process and may involve conditions such as chronic gastric mucosa inflammation, atrophic gastritis (AG), intestinal metaplasia (IM), and eventually GC. It is a complex disease arising from the interaction of environmental (*e. g.* diet, high salt intake, *etc.)* and host-associated factors (*e. g. H. Pylori* infection, genetics, age, *etc.*). In addition, GC is diagnosed two times more often in males than in females [34].

ASR (World) per 100 000

- ≥ 10.8
- 7.1–10.8
- 5.0–7.1
- 3.8–5.0
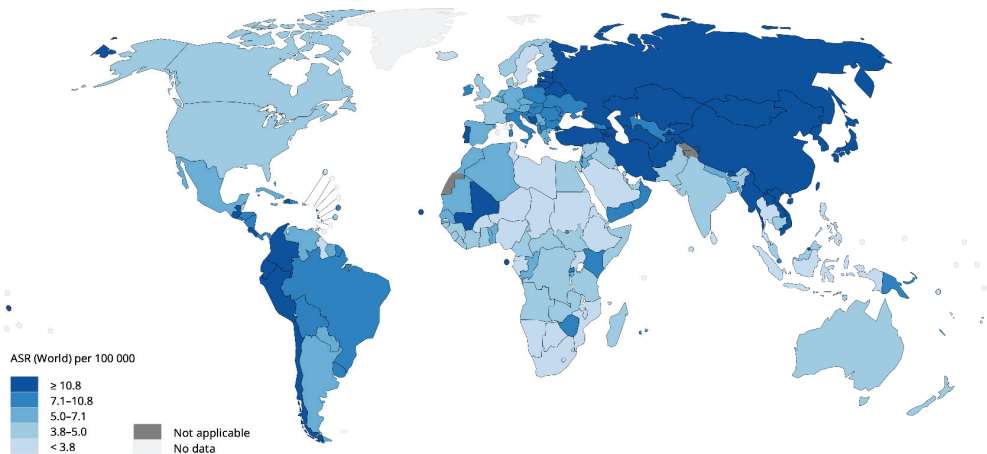- < 3.8
- Not applicable
- No data

Fig. 1.2. Estimated age-standardized incidence rates (ASR) of gastric cancer in both sexes, all ages, around the world. Data from GLOBOCAN 2020.

Only a small proportion (approx. 10 %) of GC cases are associated with family history and only 1–3 % have germline mutations – mostly in the *CDH1* gene encoding E-cadherin [35]. Other genes that may be involved in hereditary diffuse GC pathogenesis include *PALB2*, *FBXO24*, *DOT1-L* genes [36,37]. However, in most inherited cases the underlying genetic alterations are unknown. Over the past decade, studies have contributed to a more comprehensive molecular picture of GC and showed that driver alterations, including but not limited to point the mutations, are mostly detected in oncogenes such as *KRAS, PIK3CA, CTNBB1, ERBB3*, *etc*., and tumour suppressors such as *TP53, CDH1, ARID1A, FAT4, etc.* [38] (Fig. 1.3.). Other genetic changes involve somatic copy number alterations (CNA), gene fusions, structural variants causing altered splicing, epigenetic (methylation status and miRNA deregulation), and transcriptional changes [38] (Fig. 1.3.).

miRNA:
Onco: miR21, miR27a, miR130b, miR365, miR200, miR210
Tumour-supp: miR375, miR486, miR204, miR29c

Mutations:
Onco: *KRAS, PIK3CA, CTNBB1, NRG1, ERBB3, ERBB4, RHOA*
Tumour-supp: *TP53, CDH1, ARID1A, BCOR, FAT4, RNF43, MUC6*

Gene fusion:
RAF fusions, *CD44-SLC1A2*, *CLDN18-ARHGAP26*, *ROS1* fusions

Alternative splicing:
CD44, ZAK kinase, PPP1R1B-STARD3

Gastric cancer

DNA methylation:
CHD1, RUNX3, p16, hMLH1, PLA2G2A, SULF2, BMP4

Transcription:
AMPK/HNF4a/Wnt5a pathway

CNV:
Amps: *HER2, FGFR2, EGFR, METR, KRAS, NRAS, VEGFA, CCND1, CCNE1, CDK6, GATA4, KLF5, OCT1*
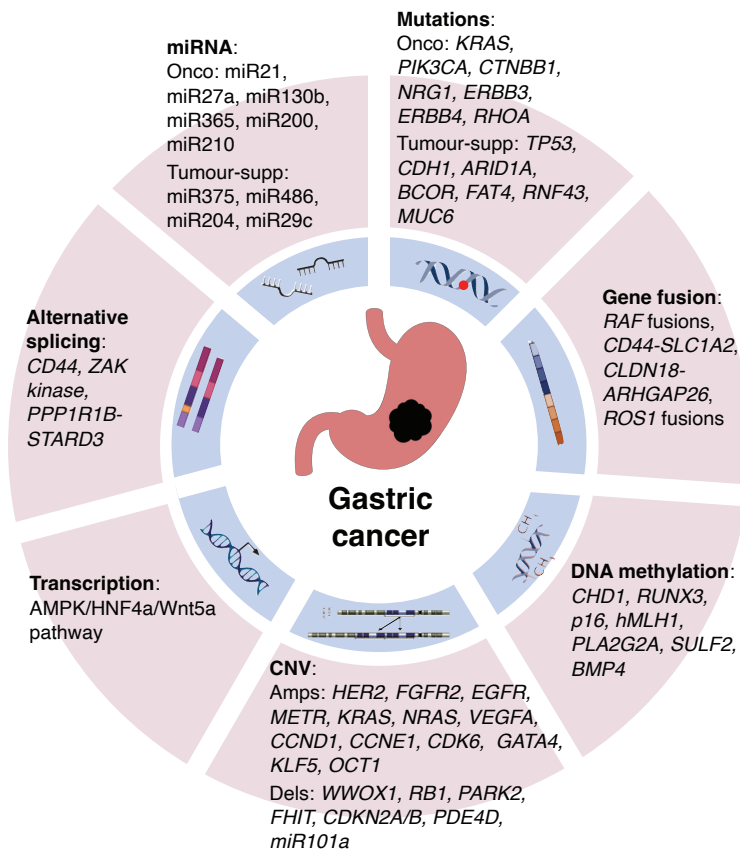Dels: *WWOX1, RB1, PARK2, FHIT, CDKN2A/B, PDE4D, miR101a*

Fig. 1.3. *Genetic and epigenetic characteristics of gastric cancer (adapted from Tan, P., and Yeoh K.G. 2015).*

Current diagnostics include histological examination after endoscopic biopsy and disease staging using computed tomography (CT), endoscopic ultrasound, positron emission tomography (PET), and laparoscopy. Despite the variety of instrumental examinations, there are still several highly important diagnostic issues: lethality that is associated with diagnosis at an advanced stage, and difficulty to detect complications such as minimal residual disease or tumour spread to the peritoneal cavity [39]. Conventional diagnostic techniques or current molecular biomarkers have a very limited role for both, detecting GC at early stages and identifying its complications. Moreover, standard needle biopsies could be the source of some minor or major complications (*e. g.* pain, bleeding, or perforation) [40]. Taking all together, minimally invasive biomarkers that would help to improve GC patients' care, diagnostics, and monitoring are highly needed. Many studies suggest that by implementing genetic [41–43] and epigenetic [44–46] molecular profiling, liquid biopsy assays could significantly contribute to the solution of many challenges associated with GC diagnostics and prognostics (*e. g.* early detection, treatment response monitoring, *etc.*) Finally, researches show progress in implementing a multi-omics approach not only for the cancer subtyping [47–49] but also for cancer diagnostics. For example, Cohen *et al.* developed a protein and cfDNA analysis-based *CancerSEEK* test that performance revealed sensitivities from 69 % to 98 % and specificity > 99 % for the detection of eight common cancer types [50]. However, there are no studies implementing epigenetic, genetic, and protein multi-analyte analysis in the case of GC.

Therefore, in the present study we investigated the potential of liquid biopsy for GC in an extensive multi-level approach: from the analysis of the most suitable biological material and impact of blood processing to comparison of plasma and tissue molecular profiles, and even functional analysis *in vitro*. We employed comprehensive molecular tools (*e. g.* small RNA (smRNA), whole-exome sequencing (WES), real-time PCR (RT-PCR), *etc.*) for complete GC and AG tissue miRNA profiling, altered miRNA validation in plasma, and plasma cfDNA mutational profile analysis. As a result, all this allowed us to implement optimized blood processing protocols, determine miRNAs that are altered during the gastritis-carcinoma sequence, show the role of two widely studied miRNAs in gastric carcinogenesis, evaluate plasma cfDNA somatic alterations, and reveal that qualitative and quantitative analysis of multi-layer molecular analytes might be a promising approach for GC disease state monitoring and prognosis.

Aim and objectives

This study aimed to analyse the plasma cell-free nucleic acids for patients with gastric cancer and to evaluate the function as well as suitability of these markers for minimally invasive diagnosis and prognosis.

Objectives:

1. To investigate the microRNA profile in the tissues of patients with atrophic gastritis and gastric cancer compared with the control group.
2. To validate the expression of tissue-associated deregulated microRNAs in the blood plasma of gastric cancer patients and evaluate the functional role of specific microRNA in gastric carcinogenesis.
3. To determine gastric tissue DNA and plasma cell-free DNA mutational profiles in gastric cancer patients, and to evaluate plasma cell-free DNA association with clinical data.
4. To evaluate the applicability of multi-layer molecular profiling for gastric cancer prognosis and disease state monitoring.

SCIENTIFIC NOVELTY

This study provides: (1) differential miRNA expression profiles in gastric premalignant (atrophic gastritis), malignant (GC), and control (CON) cases and analysis of the most deregulated miRNAs in plasma for liquid biopsy approach; (2) novel findings that reveal the important role of hsa-miR-20b-5p and hsa-miR-451a-5p in gastric carcinogenesis, which may be used to develop a beneficial strategy for future cancer therapy; (3) insights regarding the urine collection as liquid biopsy source for cfDNA analysis; (4) optimized cfNA isolation protocols and detailed description of pre-processing variables (storage, hemolysis, and isolation methods) that could affect cfNA quality; (5) comparison between two plasma cfDNA sequencing approaches for the reliable detection of low-frequency alleles; (6) analysis of concordance of GC tissue and plasma cfDNA mutational profiles and potential application of liquid biopsy for disease monitoring and survival.

Taken all together, this is the first study that have provided full miRNome profiles of pre-malignant and malignant gastric tissues in the subjects of European origin, showed functional role in gastric carcinogenesis of specific epigenetic biomarkers such as hsa-miR-20b-5p, revealed concordance between GC plasma and tissue mutational profiles by using Unique Molecular Identifier (UMI) error correction and deep sequencing, and implemented 14 multilayer molecular analytes for GC disease state discrimination and survival analysis. This extensive research provides useful observations from the processing of samples and selection of sequencing approach to the applicability of liquid biopsy for GC diagnosis and prognosis.

THE LAYOUT OF THE DISSERTATION

This thesis is prepared based on six published articles (listed in section List of scientific papers) and is composed of three major parts: (1) investigation of blood-based epigenetic signatures involves the miRNA expression profiling in gastric tissues, analysis of mostly deregulated tissue-associated miRNAs in plasma, and functional miRNA analysis in vivo and in vitro; (2) investigation of blood-based genetic signatures in detail explores the suitable source for liquid biopsy, the impact of blood processing variables, comparison of different sequencing approach and plasma cfDNA mutational profile analysis for GC diagnosis and prognosis; and (3) multi-layer molecular analysis includes analysis of multi-analyte epigenetic and genetic profiles for survival and GC state discrimination. All these studies contribute to the liquid biopsy research for GC. Publications related to studies are published in peer-reviewed journals referred in Web of Science: BioTechniques, Biopreservation and bioban-

king, International Journal of Molecular Sciences, Clinical and Translational Gastroenterology, and World Journal of Gastroenterology. The thesis design and structure are represented in Fig. 1.4.
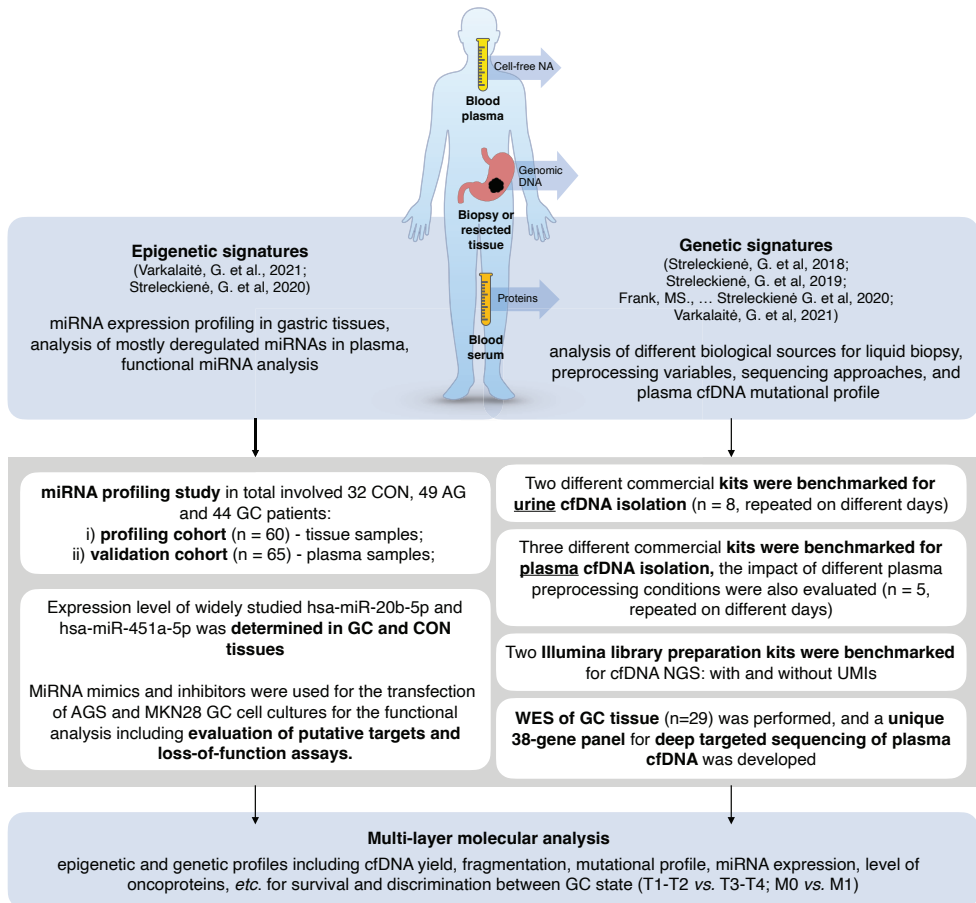

Fig. 1.4. *Structure of the thesis.*

CON – control, AG – atrophic gastritis; GC – gastric cancer; cfDNA- cell-free DNA; UMI – unique molecular identifier; WES – whole exome sequencing.

PHD CANDIDATE'S CONTRIBUTION

Contribution of author **Greta Varkalaitė** is presented below in respect to the each of publications related to the dissertation (A1-A6 listed in *List of scientific papers*).

A1: contributed to study conception and design of the research, collection of clinical and phenotypical data of study participants, performed all the investigations and laboratory experiments (total RNA and cfDNA isolation, quality control, RT-PCR), analysis of data (smRNA-seq data analysis, differential expression, and statistical analysis), and drafted the manuscript. Bioinformatic packages used are described in more detail in *1.5 Next-generation sequencing and bioinformatics data analysis* section.

A2: performed laboratory experiments (RT-PCR, Western Blot, cell culture liposome-based transfection, MTT, colony formation, apoptosis, wound healing assays), formal analysis (*in silico* miRNA target prediction, statistical analysis), and data visualisation, prepared original manuscript draft, contributed to the writing-review, and editing of article.

A3: performed cfDNA isolations (urine samples), quality control, and quantification, cfDNA sequencing library construction and sequencing experiments, analysed the data and contributed to the preparation of the manuscript.

A4: performed all laboratory experiments and investigations (blood processing, cfDNA isolation from plasma samples, quality control, and quantification, RT-PCR), statistical analysis and visualization of data, prepared original manuscript draft.

A5: was responsible for data analysis (cfDNA UMI sequencing data analysis), contributed to the preparation of the manuscript, and received full approval from the corresponding author to defend this article in the dissertation. Bioinformatic packages used are described in more detail in *1.5 Next-generation sequencing and bioinformatics data analysis* section.

A6: performed all laboratory experiments (DNA isolation from tissue samples, cfDNA isolation from plasma samples, ELISA, libraries' preparation for WES, and targeted cfDNA sequencing), collection of clinical and phenotypical data of study participants, data analysis and interpretation (WES and cfDNA UMI sequencing data analysis, gene list pathway enrichment, discriminant analysis based on multi-layer modelling, statistical analysis), visualized the data, wrote original manuscript draft. Bioinformatic packages used are described in more detail in *1.5 Next-generation sequencing and bioinformatics data analysis* section.

CO-AUTHORS' CONTRIBUTION

- **Juozas Kupčinskas** contributed to supervision and conceptualization of the studies, approved the final versions of the manuscripts, performed writing-review and editing of article A1, A2, A6; organized clinical data collection for study A2.
- **Jurgita Skiecevičienė** contributed to supervision and conceptualization of the studies, writing original draft, approved the final versions of the manuscripts, performed writing-review and editing of A1, A2, A4, A6.
- **Limas Kupčinskas** contributed to study conception and design of A1.
- **Rita Gudaitytė, Vytenis Petkevičius, Antanas Mickevičius** collected biological material and clinical data from the participants for A1.
- **Rokas Lukoševičius** contributed to laboratory work and investigations when preparing A4.
- **Gediminas Kiudelis** organized biological material and clinical data collection from the participants for A2.
- **Evelina Vaitkevičiūtė** performed laboratory work and investigations, analysed the data, and drafted the manuscript A1.
- **Simonas Juzėnas** reviewed and edited the manuscript A1; helped to develop study design, provided software and initial analysis algorithms, contributed to preparation of original A2 draft.
- **Rūta Inčiūraitė** reviewed and edited the manuscript A1; contributed to laboratory work and investigations when preparing A2 and A4; contributed to preparation of original A2 draft.
- **Violeta Šaltenienė** revised the A1 for important intellectual content; contributed to study design when preparing A2.
- **Rūta Steponaitienė, Ugnė Gyvytė** contributed to study design of A2.
- **Mārcis Leja** contributed to study conception and design of research when preparing A1; organized biological material and/or clinical data collection from the participants for A2.
- **Alexander Link** reviewed and approved the final version of the manuscript A1; organized resources for conduction of experiments related to A2.
- **Paulius Ruzgys, Sabine Franke, Cosima Thon** performed part of investigations and experiments related to A2.
- **Eugenija Kupčinskienė, Saulius Šatkauskas** provided research-related resources in Vytautas Magnus University when preparing A2.
- **Michael Forster** contributed to data analysis and helped to prepare the manuscript A3; developed A4 and A5 study design; contributed to data

collection, data analysis, interpretation, and visualization when preparing A5 and A6; performed A6 manuscript review and editing.

- **Andre Franke** contributed to supervision and conceptualization of study, performed manuscript review, and editing of A6.
- **Hayley M Reid** contributed to laboratory work and investigations, contributed to preparation of the manuscript A3.
- **Norbert Arnold, Dirk Bauerschlag** provided samples and clinical data from the participants, contributed to writing the manuscript A3.
- **Janina Fuß** designed the study, contributed to writing of original draft, organized lab work, drafted the graphical abstract of A5.
- **Malene Støchkel Frank, Julie Gehl** contributed to writing, interpreted the data when preparing A5.
- **Tim Alexander Steiert** contributed to writing, analysed data when preparing A5.

LIST OF SCIENTIFIC PAPERS

Publications related to the results of the dissertation in order of below-described results:

A1: **Varkalaitė, Greta;** Vaitkevičiūtė, Evelina; Inčiūraitė, Rūta; Šaltenienė, Violeta; Juzėnas, Simonas; Petkevičius, Vytenis; Gudaitytė, Rita; Mickevičius, Antanas; Link, Alexander; Kupčinskas, Limas; Leja, Mārcis; Kupčinskas, Juozas; Skiecevičienė, Jurgita. Atrophic gastritis and gastric cancer tissue miRNome analysis reveal hsa-miR-129-1 and hsa-miR-196a as potential early diagnostic biomarkers // *World journal of gastroenterology*: WJG. Pleasanton, CA: Baishideng Publishing Group. ISSN 1007-9327, 2022, vol. 28, no. 6, p. 653–663. [**Impact factor: 5.742**, aggregate impact factor: 6.327, quartile: Q2]

A2: **Streleckienė, Greta;** Inčiūraitė, Rūta; Juzėnas, Simonas; Šaltenienė, Violeta; Gyvytė, Ugnė; Kiudelis, Gediminas; Leja, Mārcis; Ruzgys, Paulius; Šatkauskas, Saulius; Kupčinskienė, Eugenija; Franke, Sabine; Thon, Cosima; Link, Alexander; Kupčinskas, Juozas; Skiecevičienė, Jurgita. miR-20b and miR-451a are involved in gastric carcinogenesis through the PI3K/AKT/mTOR signaling pathway: data from gastric cancer patients, cell lines and ins-gas mouse model // *International journal of molecular sciences*. Basel: MDPI. ISSN 1422–0067, 2020, vol. 21, no. 3, p. 1–16. [**Impact factor: 5.923**, aggregate impact factor: 6.387, quartile: Q1]

A3: **Streleckienė, Greta;** Reid, Hayley M; Arnold, Norbert; Bauerschlag, Dirk; Forster, Michael. Quantifying cell free DNA in urine: comparison between commercial kits, impact of gender and inter-individual variation // *BioTechniques*. New York: Biotechniques office. ISSN 0736-6205, 2018, vol. 64, no. 5, p. 225–230. [**Impact factor: 1.659**, aggregate impact factor: 3.943, quartile: Q4]

A4: **Streleckienė, Greta;** Forster, Michael; Inčiūraitė, Rūta; Lukoševičius, Rokas; Skiecevičienė, Jurgita. Effects of quantification methods, isolation kits, plasma biobanking, and hemolysis on cell-free DNA analysis in plasma // *Biopreservation and biobanking*. New Rochelle, NY: Mary Ann Liebert. ISSN 1947-5535, 2019, vol. 17, no. 6, p. 553–561. [**Impact factor: 1.906**, aggregate impact factor: 4.363, quartile: Q3]

A5: Frank, Malene Støchkel; Fuß, Janina; Steiert, Tim Alexander; **Streleckienė, Greta**; Gehl, Julie; Forster, Michael. Quantifying sequencing error and effective sequencing depth of liquid biopsy NGS with UMI error correction // *BioTechniques*. London : Future Science.

ISSN 0736-6205, 2021, vol. 70, no. 4, p. 226–-232. [**Impact factor: 1.993**, aggregate impact factor: 4.995, quartile: Q4]

A6: **Varkalaitė, Greta;** Forster, Michael; Franke, Andre; Kupčinskas, Juozas; Skiecevičienė, Jurgita. Liquid Biopsy in Gastric Cancer: Analysis of Somatic Cancer Tissue Mutations in Plasma Cell- Free DNA for Predicting Disease State and Patient Survival // *Clinical and Translational Gastroenterology*. Philadelphia: Lippincott Williams & Wilkins. ISSN 2155-384X, 2021, vol. 12, p. 1–9. [**Impact factor: 4.488**, aggregate impact factor: 6.327, quartile: Q2]

CONFERENCE PRESENTATIONS

1. Oral presentation. **G. Varkalaitė,** M. Forster, J. Kupčinskas, J. Skiecevičienė. *Plazmos laisvai cirkuliuojančios DNR analizė: mutantinio alelio dažnio dinamika skrandžio vėžio ligos eigoje (angl. Plasma cell-free DNA analysis: dynamics of mutant allele frequency during course of gastric cancer)*. Bioateitis 2021, November 25, Kaunas, Lithuania.

2. Poster presentation. **G. Varkalaitė**, E. Vaitkevičiūtė, V. Šaltenienė, S. Juzėnas, R. Inčiūraitė, J. Kupčinskas, J. Skiecevičienė. *MicroRNA signatures in atrophic gastritis and gastric cancer*. Life Science Baltics 2021, September 22–24, online.

3. Oral presentation. **G. Streleckienė.** *Skrandžio vėžiu sergančiųjų plazmos laisvai cirkuliuojančios DNR mutacijų profilio analizė, sąsajos su klinikiniais duomenimis ir pacientų išgyvenamumu* (*angl. Analysis of gastric cancer patients' plasma cell-free DNA mutational profile, correlation with clinical data and patients' survival*). Bioateitis 2020, December 4, 2020, online.

4. Oral presentation. **G. Streleckienė**, J. Arštikytė, prof. L. Kupčinskas, dr. J. Skiecevičienė. *Cell-free DNA yield and next-generation sequencing analysis in plasma of gastric cancer patients.* Health for All 2020, November 19–20, online.

5. Poster presentation. **G. Streleckienė**, M. Forster, K. Balčiūtė, J. Kupčinskas, L. Kupčinskas, J. Skiecevičienė. *Tracking the somatic gastric cancer tissue mutations in the blood: proportion of the patients with detectable alterations and association with oncoproteins*. 28th UEG Week Virtual 2020, October 11–13.

6. Poster presentation. **G. Streleckienė**, M. Forster, L. Kupčinskas, J. Skiecevičienė. *Cell-free DNA somatic alterations in plasma of gastric cancer patients: mutational spectra, association with tumor size and survival*. Virtual EHMSG (European Helicobacter & Microbiota Study Group), September 10-12, 2020.

7. Poster presentation. **G. Streleckienė**, M. Forster, L. Kupčinskas, J. Skiecevičienė. *Somatic mutation analysis of cell-free DNA in the plasma of gastric cancer patients*. Virtual ESHG (European Society of Human Genetics), June 6–9, 2020.

8. Oral presentation. **G. Streleckienė.** *Skrandžio vėžiu sergančiųjų plazmos laisvai cirkuliuojančios DNR kiekio ir mutacijų profilio analizė (angl. Analysis of the cell-free DNA yield and mutational profile in the plasma of gastric cancer patients*). Bioateitis 2019. December 11, 2019, Kaunas, Lithuania.

9. Poster presentation. **G. Streleckienė**, M. Forster, J. Arštikytė, J. Kupčinskas, L. Kupčinskas, J. Skiecevičienė. *cfDNA mutation profile analysis for gastric cancer patients*. EHMSG (European Helicobacter & Microbiota Study Group), September 9-12, 2019, Innsbruck, Austria.

10. Oral presentation. **G. Streleckienė**, M. Forster, J. Arštikytė, J. Kupčinskas, L. Kupčinskas, J. Skiecevičienė. *Cell-free DNA amount and mutation profile analysis for gastric cancer patients.* 27th UEG Week 2019, October 19–23, Barcelona, Spain.

11. Oral presentation. **G. Streleckienė**, M. Forster, J. Kupčinskas, L. Kupčinskas, J. Skiecevičienė. *cfDNA amount and mutation profile analysis for gastric cancer patients*. International Doctoral and Resident Students Conference: Science for Health 2019, April 9, Kaunas, Lithuania.

12. Poster presentation. **G. Streleckienė**, J. Kupčinskas, S. Juzėnas, V. Šaltenienė, J. Skiecevičienė, G. Kiudelis, L. Jonaitis, R. Lukoševičius, R. Inčiūraitė, P. Malfertheiner, L. Kupčinskas, A. Link, J. Skiecevičienė. *Inhibition of miR-20b reduces cell viability and colony formation in gastric cancer by targeting phosphatase and tensin homolog (PTEN) and thioredoxin-interacting protein (TXNIP) genes*. 26th UEG Week 2018, October 20 -24, Vienna, Austria.

13. Oral presentation. **G. Streleckienė**, J. Kupčinskas, S. Juzėnas, V. Šaltenienė, G. Kiudelis, L. Jonaitis, J. Arštikytė, M. Leja, C. Langner, A. Franke, J. Skiecevičienė, P. Malfertheiner, A. Link, L. Kupčinskas. *Inhibition of miR-20b reduces cell viability and colony formation in gastric cancer by targeting phosphatase and tensin homolog (PTEN) and thioredoxin-interacting protein (TXNIP) genes.* EHMSG XXXIth International Workshop, September 14–15, 2018, Kaunas, Lithuania.

14. Oral presentation. **G. Streleckienė**. *miR-20b acts as potential oncogenic miRNA in gastric cancer derived cell lines by targeting phosphatase and tensin homolog (PTEN) and thioredoxin-interacting protein (TXNIP) genes*. XVth International Conference of the LBS (Lithuanian Biochemical Society), June 26-28, 2018, Dubingiai, Lithuania.

15. Oral presentation. **G. Streleckienė**, S. Juzėnas, V. Šaltenienė, J. Kupčinskas, J. Skiecevičienė, L. Kupčinskas. *Inhibition of miR-20b reduces cell viability and colony formation in gastric cancer potentially by targeting Phosphatase and Tensin Homolog (PTEN) gene*. Science for Health 2018, April 13, 2018, Kaunas, Lithuania.

16. Oral presentation. **G. Streleckienė**. *Hsa-miR-20b-5p, hsa-miR-451a-5p ir hsa-miR-1468- 5p vaidmuo skrandžio vėžio patogenezėje (angl. Role of hsa-miR-20b-5p, hsa-miR-451a-5p and hsa-miR-1468-5p in the*

pathogenesis of gastric cancer). Bioateitis 2017, December 7, 2017, Vilnius, Lithuania.

17. Oral presentation. **G. Streleckienė**, J. Kupčinskas, S. Juzėnas, R. Gedgaudas, V. Šaltenienė, A. Link, M. Leja, G. Kiudelis, L. Jonaitis, R. Lukoševičius, R. Inčiūraitė, G. Hemmrich-Stanisak, M. Hübenthal, A. Franke, P. Malfertheiner, L. Kupčinskas, J. Skiecevičienė. *Role of miR-20b-5p, miR-451a-5p and miR-1468-5p in gastric cancer*. 2017 Fit for the Future in GI, November 17 to November 18, 2017, Berlin, Germany.

18. Oral presentation. **G. Streleckienė**, J. Kupčinskas, S. Juzėnas, R. Gedgaudas, V. Šaltenienė, A. Link, M. Leja, G. Kiudelis, L. Jonaitis, R. Lukoševičius, R. Inčiūraitė, G. Hemmrich-Stanisak, M. Hübenthal, A. Franke, P. Malfertheiner, L. Kupčinskas, J. Skiecevičienė. *Role of miR-20b-5p, miR-451a-5p and miR-1468-5p in gastric cancer*. 25th UEG Week 2017, October 28 to November 1, 2017, Barcelona, Spain.

1. SUMMARY OF MATERIALS AND METHODS

1.1. Ethics statement

Studies involved in the dissertation were approved by Kaunas Regional Biomedical Research Ethics (No. BE-2-10, No. BE-2-31) and University Hospital of Schleswig-Holstein's Ethics (No. B327/10, No. D470/14) Committees (Supplements 1–2 and 3–4). Principles of the Helsinki Declaration were followed, and all subjects provided written informed consent.

1.2. Study population

In total, miRNome profiling study **("Epigenetic signatures")** involved 32 CON, 49 AG, and 44 GC patients, who were divided into (1) profiling (n = 60) and (2) validation cohorts (n = 65). The profiling cohort consisted of gastric samples, and validation cohort – plasma samples (including available samples from study part *1.2.2 cfDNA mutational profiling study*). CON group consisted of subjects, who had no signs of gastric mucosal atrophy or intestinal metaplasia (stage 0 according to Operative Link on Gastritis Assessment (OLGA) classification). AG group consisted of individuals that had stage I-IV atrophy score in gastric mucosa by OLGA classification.

Detailed clinical characteristics are presented in the article "*Atrophic gastritis and gastric cancer tissue miRNome analysis reveal hsa-miR-129-1 and hsa-miR-196a as potential early diagnostic biomarkers*".

Expression analysis of hsa-miR-20b-5p and hsa-miR-451a-5p in functional miRNA study **("Epigenetic signatures"**) was performed on gastric tissue biopsy samples from 13 GC patients. Gastric tissue samples of CON group (n = 13) were collected from healthy subjects without atrophy or intestinal metaplasia.

Functional tests were performed using commercially available cell cultures AGS and MKN28, developed from primary and metastatic GC sites, respectively.

The characteristics of patients and detailed cell cultivation description are presented in the article "*miR-20b and miR-451a Are Involved in Gastric Carcinogenesis through the PI3K/AKT/mTOR Signaling Pathway: Data from Gastric Cancer Patients, Cell Lines and Ins-Gas Mouse Model*".

Benchmarking studies **("Genetic signatures")** of the urine (n = 8) and plasma (n = 5) cfDNA analysis involved self-reported healthy volunteers. The biological material collection was repeated on separate days (2–5 depending on the experiment), samples were centrifuged according to sample processing guidelines and pooled in equal volumes.

Characteristics and detailed descriptions are presented in the articles: "*Quantifying cell free DNA in urine: comparison between commercial kits, impact of gender and inter-individual variation*" and "*Effects of Quantification Methods, Isolation Kits, Plasma Biobanking, and Hemolysis on Cell-Free DNA Analysis in Plasma*".

Concordance analysis between GC tissue and plasma somatic mutational profiles in study of cfDNA mutation analysis **("Genetic signatures")** were performed for treatment-naive GC patients (n = 29). Paired tissue and plasma samples were collected at the same time point for the same patient. CON group (n = 20) consisted of self-reported healthy subjects without a history of cancer.

Detailed clinical characteristics are presented in the article "*Liquid biopsy in gastric cancer: analysis of somatic cancer tissue mutations in plasma cell-free DNA for predicting disease state and patient survival*".

1.3. Isolation and quantification of nucleic acids

Genomic DNA (gDNA) from the primary GC tissues and white blood cells (WBCs) were isolated using the AllPrep DNA/RNA Mini Kit (Qiagen) and salting-out methods respectively. Total RNA from CON, AG and GC tissues was extracted using miRNeasy Mini Kit or miRNeasy Micro Kit (Qiagen). The quantity and quality of the isolated samples were assessed using the following systems: Nanodrop2000 spectrophotometer (Thermo Fisher Scientific), and/or automated capillary electrophoresis TapeStation 2000 system (Agilent Technologies), and/or automated capillary electrophoresis Agilent 2100 Bioanalyzer system (Agilent Technologies).

Two commercially available kits were benchmarked for urine samples: (1) NEXTprep-Mag Urine cfDNA Isolation Kit (PerkinElmer); (2) Urine Cell-Free Circulating DNA Purification Midi Kit (Norgen Biotek). Furthermore, performance of three commercially available isolation kits for plasma samples was analysed: (1) The NextPrep-Mag cfDNA Isolation Kit (Bioo Scientific); (2) MagMAX Cell-Free DNA Isolation Kit (Thermo Fisher Scientific); (3) QIAamp Circulating Nucleic Acid Kit (Qiagen). The quantity and length distribution of cfDNA was evaluated by TapeStation 2200 system (Agilent Technologies) using the high sensitivity screen tapes and reagents.

All procedures were handled according to the manufacturer's instructions. Detailed descriptions are presented in articles: "*Quantifying cell free DNA in urine: comparison between commercial kits, impact of gender and inter-individual variation*" and "*Effects of Quantification Methods, Isolation Kits, Plasma Biobanking, and Hemolysis on Cell-Free DNA Analysis in Plasma*".

1.4. Expression analysis of miRNA by quantitative real-time PCR

Quantitative RT-PCR was performed for the evaluation of specific microR-NAs' expression, and absolute hsa-miR-223-3p quantification. Following TaqMan MicroRNA Assays were used: hsa-miR-129* (Assay ID: 002298), hsa-miR-196a (Assay ID: 241070_mat), has-miR-223 (Assay ID: 002098), hsa-miR-20b-5p (Assay ID: 001014) and hsa-miR-451a-5p (Assay ID: 001141) on 7500 Fast Real-Time PCR System (Applied Biosystems). Cycle threshold (Ct) values were normalized to the RNU6B (Assay ID: 0010930) or has-miR-16 (Assay ID: 000391) (Thermo Fisher Scientific) endogenous control.

A detailed description is presented in the articles "*Atrophic gastritis and gastric cancer tissue miRNome analysis reveal hsa-miR-129-1 and hsa-miR-196a as potential early diagnostic biomarkers*", "*miR-20b and miR-451a Are Involved in Gastric Carcinogenesis through the PI3K/AKT/mTOR Signaling Pathway: Data from Gastric Cancer Patients, Cell Lines and Ins-Gas Mouse Model*", and "*Effects of Quantification Methods, Isolation Kits, Plasma Bio-banking, and Hemolysis on Cell-Free DNA Analysis in Plasma*".

1.5. Next-generation sequencing and bioinformatics data analysis

Small RNA sequencing (smRNA-seq) libraries were prepared using Illumina TruSeq Small RNA Sample Preparation Kit (Illumina) and Illumina HiSeq 2500 platform according to the manufacturer's instructions. Analysis of raw small RNA sequencing data was performed by nf-core/smrnaseq pipeline v.1.0.0 (miRBase v.22.1 [29]). Detailed description regarding the preparation of NGS libraries and analysis workflow is presented in the paper "*Atrophic gastritis and gastric cancer tissue miRNome analysis reveal hsa-miR-129-1 and hsa-miR-196a as potential early diagnostic biomarkers*".

Exome sequencing was performed for GC tissue and patient-matched WBCs samples. Sequencing libraries were constructed using the Illumina TruSeq Nano DNA Library Prep Kit (Illumina) and captured using the Integrated DNA Technologies xGen Exome Research Panel (Integrated DNA Technologies). Deep sequencing (40 000 × raw coverage) was performed for plasma cfDNA. Libraries were constructed using TruSight Oncology UMI Reagents (Illumina) and captured using Integrated DNA Technologies xGen Custom Panel (Integrated DNA Technologies). Samples were sequenced on the NovaSeq 6000 platform (Illumina) according to the manufacturer's instructions. The GATK Best Practices paired-sample workflow [51] was used for the exome sequencing data analysis (human genome reference build hg19). Plasma cfDNA sequencing analysis was performed using the Illumi-

na UMI Error Correction App (v1.0.0.1). Detailed description regarding the analysis workflow is presented in the paper "*Liquid biopsy in gastric cancer: analysis of somatic cancer tissue mutations in plasma cell-free DNA for predicting disease state and patient survival*".

1.6. Functional miRNA analysis

Synthetic miRNAs, hsa-miR-451a-5p and hsa-miR-20b-5p, anti-hsa-miR-20b-5p, and non-specific miRNA negative controls (miRVana, Ambion by Thermo Fisher Scientific) were used for the transfection of AGS and MKN28 cell lines for the functional analysis. This analysis included the evaluation of putative targets and loss-of-function assays. Analysis of putative targets was performed in the following order: (1) *in silico* analysis (*DIANA Lab Tools TarBase, miRanda, TargetScan*), (2) Western Blot, and (3) luciferase reporter assay. Cell culture assays performed in the study were the following: 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) assay for viability and proliferation, clonogenic assays for colony formation ability, flow cytometry FITC Annexin V and PI assay for apoptosis, and wound healing assay for cell migration. Full description of the procedures is presented in the article "*miR-20b and miR-451a Are Involved in Gastric Carcinogenesis through the PI3K/AKT/mTOR Signaling Pathway: Data from Gastric Cancer Patients, Cell Lines and Ins-Gas Mouse Model*".

1.7. Protein analysis

Potential target-gene analysis of hsa-miR-20b-5p and hsa-miR-451a-5p was performed by evaluating protein expression levels by Western Blot (WB). Total protein from transfected GC commercial cells was extracted by using 1× radioimmunoprecipitation assay (RIPA) buffer (Abcam) (Sigma Aldrich). Pierce BCA Protein Assay Kit (Thermo Scientific) was used for the total protein evaluation. Following antibodies were used in the study: IRF1 (ab186384; Abcam), PTEN (ab32199, Abcam), TXNIP (40-3700; Thermo Fisher Scientific), CAV1 (ab192869, Abcam), TSC1 (37-0400, Thermo Fisher Scientific), and GAPDH (AM4300; Ambion by Thermo Fisher Scientific). Detailed description regarding WB procedure is presented in the publication "*miR-20b and miR-451a Are Involved in Gastric Carcinogenesis through the PI3K/AKT/mTOR Signaling Pathway: Data from Gastric Cancer Patients, Cell Lines and Ins-Gas Mouse Model*".

Enzyme-linked immunosorbent assay (ELISA) was used for the detection of serum oncoproteins. Following ELISA kits were used in the study: Human Carcinoembryonic Antigen (CEA) ELISA Kit (ab99992; Abcam), Human

Cancer Antigen CA 19-9 ELISA Kit (ab108642; Abcam), and Human Cancer Antigen 72-4 (Tumor Marker CA724) ELISA Kit (E-EL-H0613; Elabscience). This part of the thesis is described in the publication "*Liquid biopsy in gastric cancer: analysis of somatic cancer tissue mutations in plasma cell-free DNA for predicting disease state and patient survival*".

2. SUMMARY OF RESULTS

In this chapter, the most relevant results obtained during the period of PhD and published in dissertation-related manuscripts are described. The first part describes the results of blood-based **epigenetic** signatures study, which involved comprehensive miRNA analysis. The second part defines the results of blood-based **genetic** signatures study, which consisted of cfDNA analysis and evaluation of tissue and plasma mutational profiles. The third part – study of **multi-layer** molecular analysis – outlined blood-based molecular analytes with the highest potential to be applicable for disease state evaluation.

2.1. Epigenetic signatures

2.1.1. Analysis of tissue miRNome profiles

To determine the tissue miRNome profiles, smRNA-seq was performed and included gastric biopsy samples of CON, AG, and GC (paired cancerous and adjacent) cases. In total 1037 miRNAs annotated in the miRBase v22.1 were identified. The abundance of deregulated tissue miRNAs was associated with the control-gastritis-cancer pathological sequence: (1) 20 miRNAs showed altered expression comparing AG and CON groups; (2) 99 – comparing GC and AG, and (3) 129 – comparing GC and CON groups. (Fig. 2.1. A). The top five most deregulated miRNAs in each case are presented in Fig. 2.1. A. Multidimensional scaling analysis (MDS) of normalized miRNA expression levels revealed 4 clusters, corresponding to the study groups described above with the highest overlap between CON, GC adjacent, and AG, and a separate cluster of GC samples (Fig. 2.1. B).

Tissue smRNA-seq data analysis showed that miRNAs hsa-miR-129-1-3p and hsa-miR-196a-5p were significantly deregulated comparing all three study groups: CON, AG, and GC, and reflected a stepwise process of gastric carcinogenesis (Fig. 2.2. A). In addition to this, Spearman's correlation coefficient showed strong correlation with the CON-AG-GC sequence: -0.7561 ($p = 0.01$) and 0.7125 ($p = 1.17 \times 10^{-16}$) hsa-miR-129-1-3p and hsa-miR-196a-5p, respectively. Thus, expression changes of hsa-miR-129-1-3p and hsa-miR-196a-5p were evaluated by RT-PCR in plasma samples of independent CON, AG, GC cohorts to analyse its possible applicability as a minimally invasive biomarker. Results showed a similar expression pattern, however, only hsa-miR-129-1-3p was significantly down-regulated when comparing AG and GC groups ($p = 0.024$) (Fig. 2.2. B).
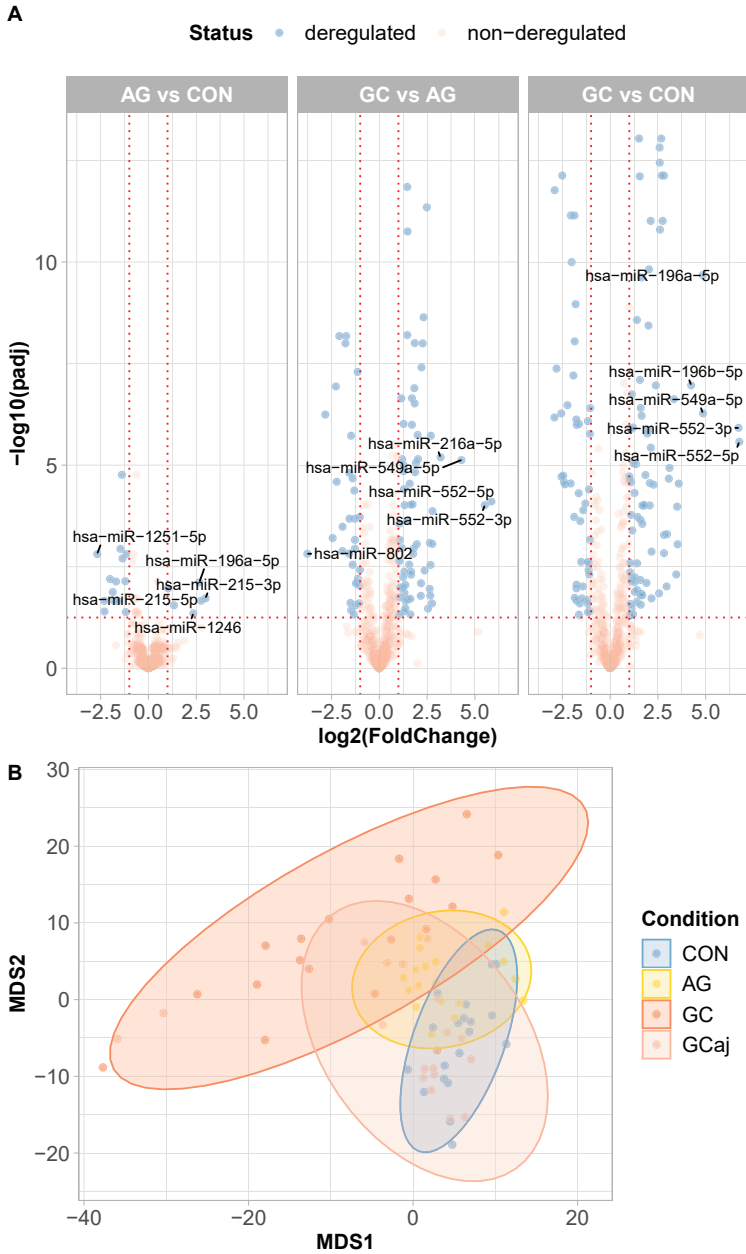
Fig. 2.1. *Results of miRNA differential expression analysis in gastric biopsy samples. **A)** Differentially expressed gastric tissue microRNAs comparing different study groups. P-adjusted < 0.05 and |log2FC| > 1; **B)** MDS plot based on normalized data showing a clustering corresponding to control (CON), atrophic gastritis (AG), gastric cancer (GC), and adjacent gastric cancer (GCaj) tissues.*
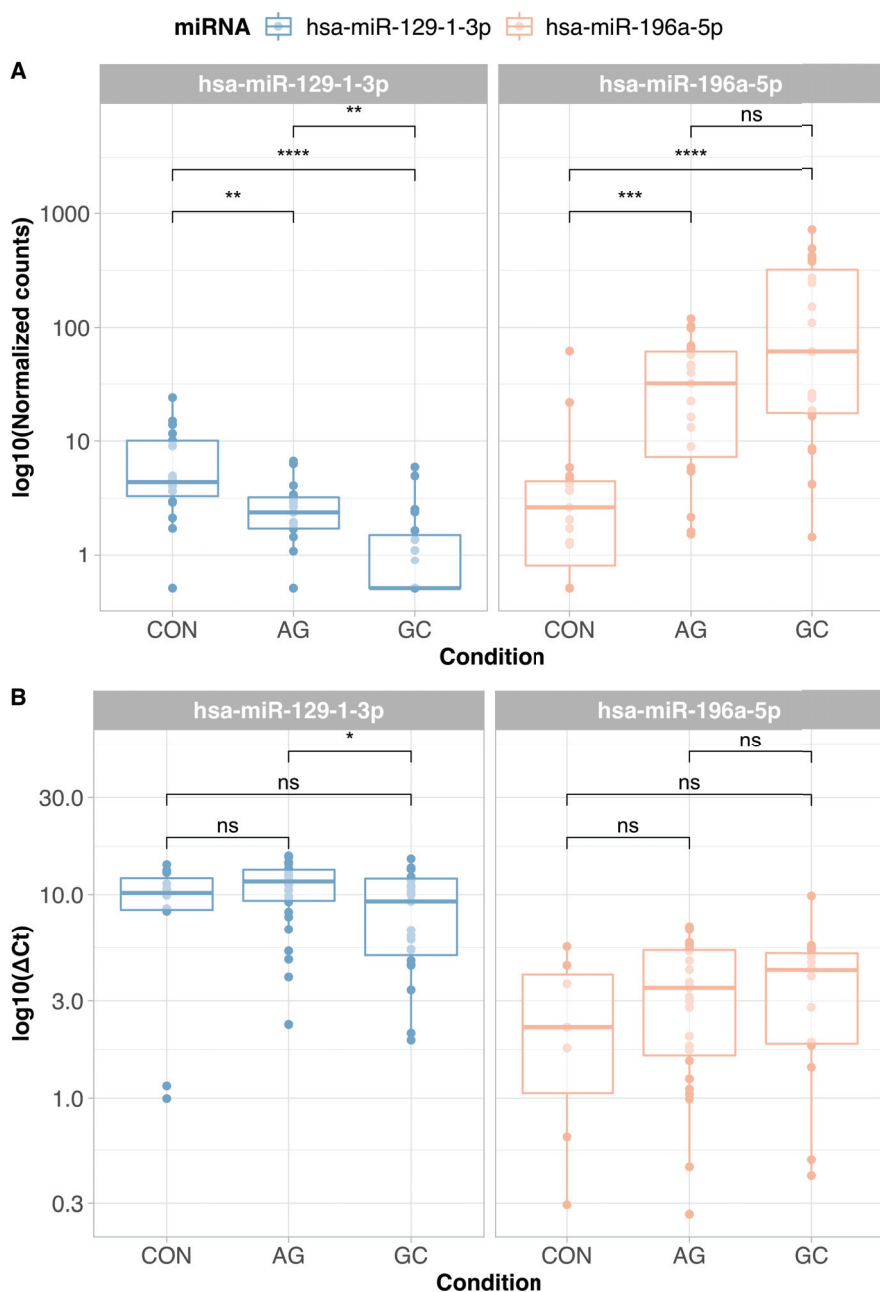
Fig. 2.2. *hsa-miR-129-1-3p and hsa-miR-196a-5p expression levels in different study groups. **A)** gastric tissue samples; **B)** plasma samples. Box plot graphs: boxes correspond to the median value and interquartile range; CON – control; AG – atrophic gastritis; GC – gastric cancer; ns – not significant; * p < 0.05; ** p < 0.01; *** p < 0.001; **** p < 0.0001.*

For a more detailed description of the results and additional figures regarding tissue miRNA analysis please refer to the article "*Atrophic gastritis and gastric cancer tissue miRNome analysis reveal hsa-miR-129-1 and hsa-miR-196a as potential early diagnostic biomarkers*".

2.1.2. Functional role of epigenetic signatures in the gastric carcinogenesis

The study of a comprehensive analysis of epigenetic signatures also involved miRNA functional role evaluation. Previous studies showed that specific miRNA families (miR-17~92 cluster (of which hsa-miR-20b-5p is a member)) or miRNAs (hsa-miR-451a-5p) are widely reported as potential oncogenes and are deregulated in many cancers [52–58]. Therefore, we have aimed to verify the expression changes of hsa-miR-20b-5p and hsa-miR-451a-5p in GC biopsies. A significant up-regulation of hsa-miR-20b-5p (p = 0.026) and down-regulation of hsa-miR-451a-5p (p = 0.039) was determined in GC comparing to CON. Next, a pilot smRNA sequencing study of several GC patients' plasma samples was performed. Analysis revealed that hsa-miR-20b-5p expression in plasma was not supported by any of the smRNA sequencing reads (data not shown) and this miRNA was not further investigated. On the other hand, the expression of hsa-miR-451a-5p was detected in plasma and supported by several hundreds of raw smRNA-seq reads (data not shown). These findings were in concordance with previously reported results by other authors [59,60]. Finally, functional role of both hsa-miR-20b-5p and hsa-miR-451a-5p was further investigated in gastric carcinogenesis. By using cell culture assays, we have, first, determined new putative target genes involved in PI3K/AKT/mTOR signalling pathway: (1) transfection of anti-hsa-miR-20b-5p significantly increased expression of PTEN (p = 0.011) and TXNIP proteins (p = 0.025); and (2) transfection of hsa-miR-451a-5p reduced the yield of CAV1 and TSC1 proteins (p = 0.011 and p = 0.024, respectively). In addition, miRNA-mRNA interaction was confirmed by luciferase reporter assay. Second, loss-of-function assays showed that anti-hsa-miR-20b-5p: reduced cell viability (by 22.1 %; p = 0.029), and increased percentage of apoptotic cells (by 12.5 %; p = 0.040). Both miRNAs dramatically reduced the number of colonies (hsa-miR-20b-5p: by 73.8 %; p = 2×10^{-4} and 60.1 %; p = 0.021, in AGS and MKN28, respectively; hsa-miR-451a-5p: by 50.6 %; p = 0.043). These findings suggest hsa-miR-20b-5p has potential oncogenic, and hsa-miR-451a-5p – tumour suppressive role in GC.

For a detailed description of the results regarding the miRNA functional analysis please refer to the article "*miR-20b and miR-451a Are Involved in Gastric Carcinogenesis through the PI3K/AKT/mTOR Signaling Pathway: Data from Gastric Cancer Patients, Cell Lines and Ins-Gas Mouse Model*".

2.2. Genetic signatures

2.2.1. Evaluation of biological material impact on cfDNA analysis

The collection of blood plasma and urine could be named as one of the most standardized and routine laboratory procedures. For this reason, plasma and urine were selected as biological materials for the benchmarks of commercially available cfDNA isolation kits. Urine cfDNA isolation analysis showed significant differences between kits: greater yield of cfDNA was isolated by NextPrep-Mag Urine (on average 2.04 ng/mL of urine) compared to Urine Cell-Free Circulating DNA isolation kit (0.83 ng/mL of urine) (p = 0.008). However, very high inter-, intra-individual, gender-based variety and DNA fractions longer than 165 nucleotides were observed.

An investigation involving plasma samples, first, revealed the importance of the quantification method. Depending on the quantification method, cfDNA yield analysis between isolation kits showed slightly different results. Second, the importance of the chosen extraction method. The yield of cfDNA isolated by QIAamp was significantly greater (on average 9.49 ng/mL of plasma) compared with the NextPrep-Mag (5.74 ng/mL of plasma), and MagMAX (6.06 ng/mL of plasma) (p = 1.62×10^{-7}; p = 1.16×10^{-6}, respectively) (fluoresce-based quantification). Furthermore, quantitative miR-223 analysis was performed to evaluate the possibility of simultaneous miRNA analysis. Higher yields of miRNA were identified in samples isolated by QIAamp and NextPrep-Mag compared to MagMAX (2×10^{-3} and 7×10^{-3} pg/mL of plasma; compared to 1×10^{-4} pg/mL of plasma; p = 0.003 and p = 0.027, respectively).

Based on the above-described results plasma was selected as a more consistent and reliable biological material for cfNA analysis. Moreover, higher cfDNA and miRNA yields isolated by the QIAamp isolation kit were determinants, therefore, this kit was selected for further studies. A more detailed description of the urine and plasma cfDNA analysis involving results of other interfering factors are presented in the respective articles "*Quantifying cell free DNA in urine: comparison between commercial kits, impact of gender and inter-individual variation*" and "*Effects of Quantification Methods, Isolation Kits, Plasma Biobanking, and Hemolysis on Cell-Free DNA Analysis in Plasma*".

2.2.2. Analysis of sequencing approach impact on cfDNA

To completely understand the cfDNA analysis process, we have estimated the possible effect of sequencing approach: signal noise (the allele frequency (AF) of PCR and sequencing errors) was evaluated using the Multiplex I cfDNA Reference Standard Set. The set included 100 % Multiplex I Wild Type

cfDNA Reference Standard, and the 0.1 % Multiplex I cfDNA Reference Standard with mutated *PIK3CA* p.E545K of 0.13 % AF. The study compared the sequencing libraries without UMIs, and with UMIs.
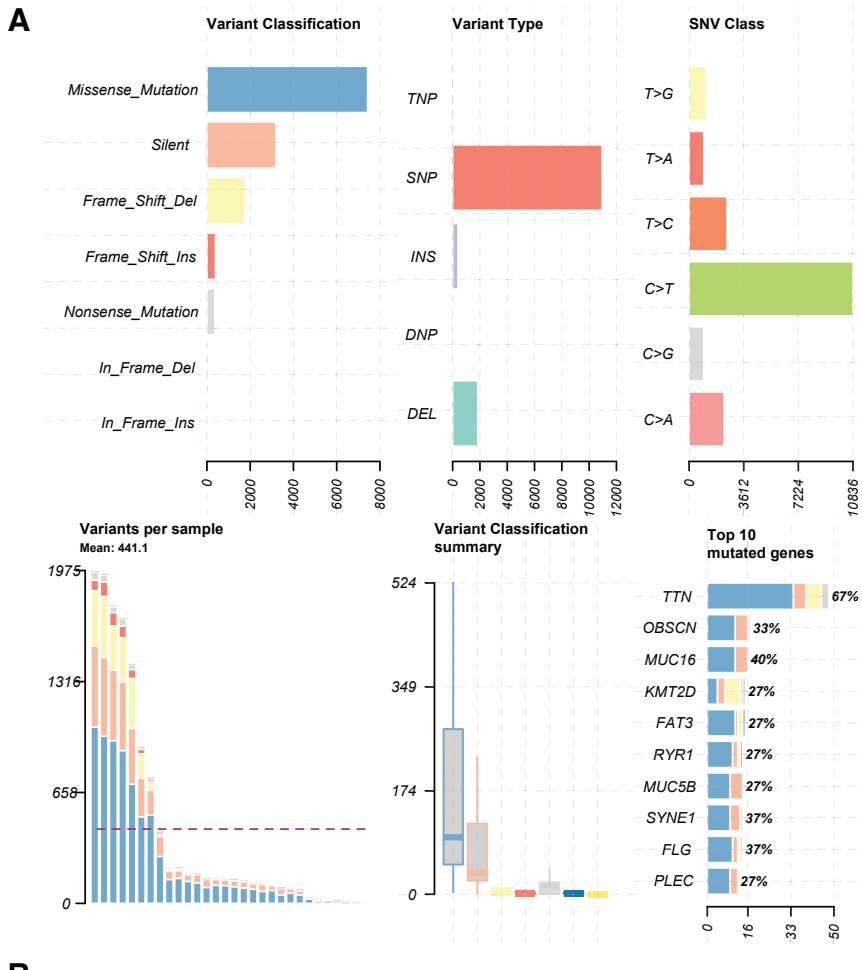
Results revealed that signal noise was lower than 0.07 % for the non-UMI libraries in 95 % of the considered genomic positions, and 0.00 % for the UMI libraries in 95 % of positions. Taking into consideration sequencing depth, stringent bioinformatic filtering of non-UMI libraries reduced the depths and thereby increased the rate of the PCR and sequencing errors. On the other hand, bioinformatic error correction of the UMI libraries dramatically reduced the effective sequencing depth but lowered the allele frequency of the signal noise. Based on the results and guidelines suggested by other researchers [61,62] working with low biomass samples, we have used very deep sequencing and barcoded each cfDNA molecule using UMIs.

The results are described in more detail in the article *"Quantifying sequencing error and effective sequencing depth of liquid biopsy NGS with UMI error correction"*.

2.2.3. Mutation analysis of cfDNA in the GC patients' plasma

After selection of the most suitable biological material, isolation kit, and sequencing approach, we have analysed total cfDNA yield determined by automated capillary electrophoresis system. A significantly higher yield of cfDNA was detected in GC patients' plasma (87.59 ng/mL of plasma) compared to CON (2.01 ng/mL of plasma) ($p = 7.07 \times 10^{-14}$). Moreover, total cfDNA yield in plasma positively correlated with serum CEA, level (R = 0.44, p = 0.02), and cfDNA fragments in GC patients' plasma were shorter compared with CON group: (1) mono-nucleosomal: 73 *vs.* 125 bp ($p = 1.60 \times 10^{-8}$), and (2) di-nucleosomal: 349 *vs.* 259 bp ($p = 1.06 \times 10^{-7}$).

To develop a custom gene panel for deep targeted sequencing of plasma cfDNA, we evaluated GC tissue mutational profile. WES for paired GC tissue and WBC (as normal control) was performed. Somatic alterations in tissue were detected for 23 of 29 patients with GC (79.31 %). The most frequent variant classes, types, and top mutated genes are presented in Fig. 2.3. A. Based on WES results unique 38-gene panel was designed (Fig. 2.3. B) and included most mutated genes that are also present in the Catalogue of Somatic Mutations in Cancer (COSMIC). Pathway enrichment analysis showed that the most mutated genes (n = 26), included in our panel, were associated with Wnt signalling (16.7 %), Cadherin signalling (14.6 %), p53 pathway feedback loops 2 (12.5 %), angiogenesis (12.5 %), EGF receptor signalling (10.4 %), CCKR signalling map (10.4 %), PI3 kinase (8.3 %), p53 (8.3 %), and Ras (6.3 %) pathways. Twelve genes from a 38 gene list were unclassified (31.6 %) and were not assigned to any PANTHER pathway [63].

Fig. 2.3. *Summary of mutational spectra (whole exome sequencing data) in GC tissue samples. **A)** absolute variant class values; absolute variant type values; distribution of various SNV substitutions; absolute numbers of variants per sample, the dashed line shows the mean quantity of somatic variants per sample; mean distribution of variant classes per sample; top 10 mutated genes, x axis: absolute numbers (samples), percentages calculated from all somatic variants detected; **B)** List of genes involved in our custom gene panel; Color codes in graphs are the same. DEL – deletion; INS – insertion; SNP – single nucleotide polymorphism; SNV – single nucleotide variant.*

Further analysis was performed in the frame of our custom panel and included filtered variants (without germline and passing Mutect2 quality control filters), and three following groups: (1) alterations detected in tissue samples; (2) alterations detected in plasma samples, and (3) alterations that overlapped between tissue and plasma samples.

I. Samples with detectable **tissue** mutations on average had 8.4 variants (from 1 to 23). Variant allele frequencies (VAFs) varied from 2.8 % to 87.1 %. Top three variant classes identified were: missense mutations – 58.6 %, silent 23.3 %, and frameshift deletions – 17.3 %. The most frequently mutated genes in GC tissue are presented in Fig. 2.4. A.

II. Somatic **plasma** cfDNA variants were detected for 21 out of 23 (91.3 %) GC patients (part I). The range of somatic variants was from 1 to 14 (5.4 on average) and VAFs ranged from 0.3% to 33.7 %. Top three variant classes identified were: intron – 57.9 %, missense – 14.3 %, and silent – 9.02 %. The most frequently mutated genes in GC plasma samples are presented in Fig. 2.4. B.

III. The **overlap** of mutations between tissue and plasma samples was observed for 11 out of 23 (47.8 %) GC patients (part I). The range of tumour-derived plasma cfDNA variants was from 1 to 12 (3.5 on average). VAFs ranged from 0.3 % to 19.4 %. Top three variant classes identified were: intron – 14.3 %, missense – 8.27 %, and silent – 5.26 %. The number and concordance of detected variants (in plasma cfDNA and tissue) are presented in Fig. 2.4. C and Venn diagram (Fig. 2. 4. D).

Association analysis of tumour-derived plasma cfDNA somatic variants with GC clinical characteristics, total cfDNA yield, serum level of oncoproteins, and age revealed that the quantity of tissue and/or plasma somatic variants showed a positive moderate correlation with age (R = 0.47, p = 0.012; R = 0.4, p = 0.035; tissue and plasma, respectively; tumour-derived cfDNA variants: R = 0.38, p = 0.04). Also, cfDNA sequence alterations that are overlapping with tumour tissue mutational profile were significantly more often observed in samples of the patients with larger tumours (T3-T4 – 55.6 % and T1-T2 – 10.0 %, p = 0.018).

Detailed description of the results is presented in the article *"Liquid biopsy in gastric cancer: analysis of somatic cancer tissue mutations in plasma cell-free DNA for predicting disease state and patient survival"*.
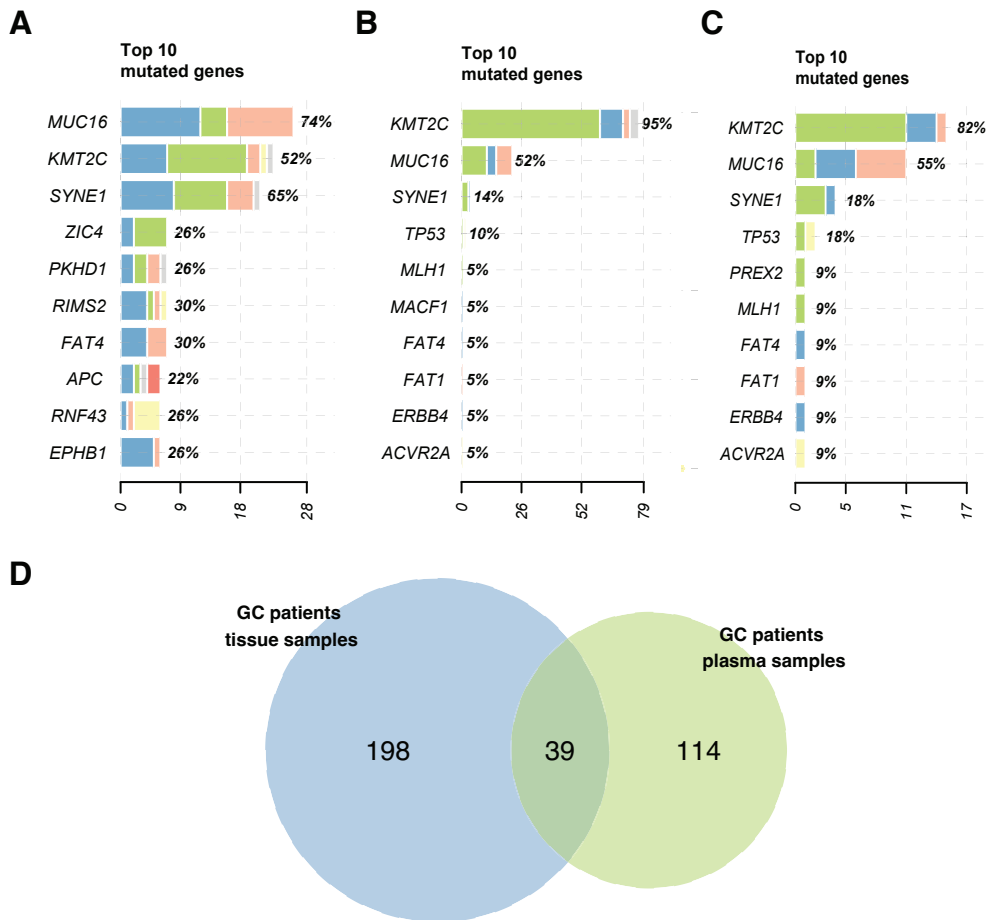
Fig. 2.4. *Top 10 most mutated genes in **A)** tissue samples; **B)** in plasma samples; **C)** overlapping variants; x axis: absolute numbers (samples), percentages calculated from all somatic variants detected; and **D)** Venn diagram showing the quantity of unique and shared somatic alterations detected in GC patient tissue and plasma samples; Colours in A-C: blue – missense; green – intron; salmon – silent; grey – nonsense; light yellow – frame shift deletion; pink – frame shift insertion.*

2.3. Multi-layer molecular modelling

Finally, we summarised all collected data to further investigate multi-layer molecular profiling applicability for GC prognosis (survival analysis) and disease state monitoring (discriminative analysis). Analytes included in the study are represented in Fig. 2.5.
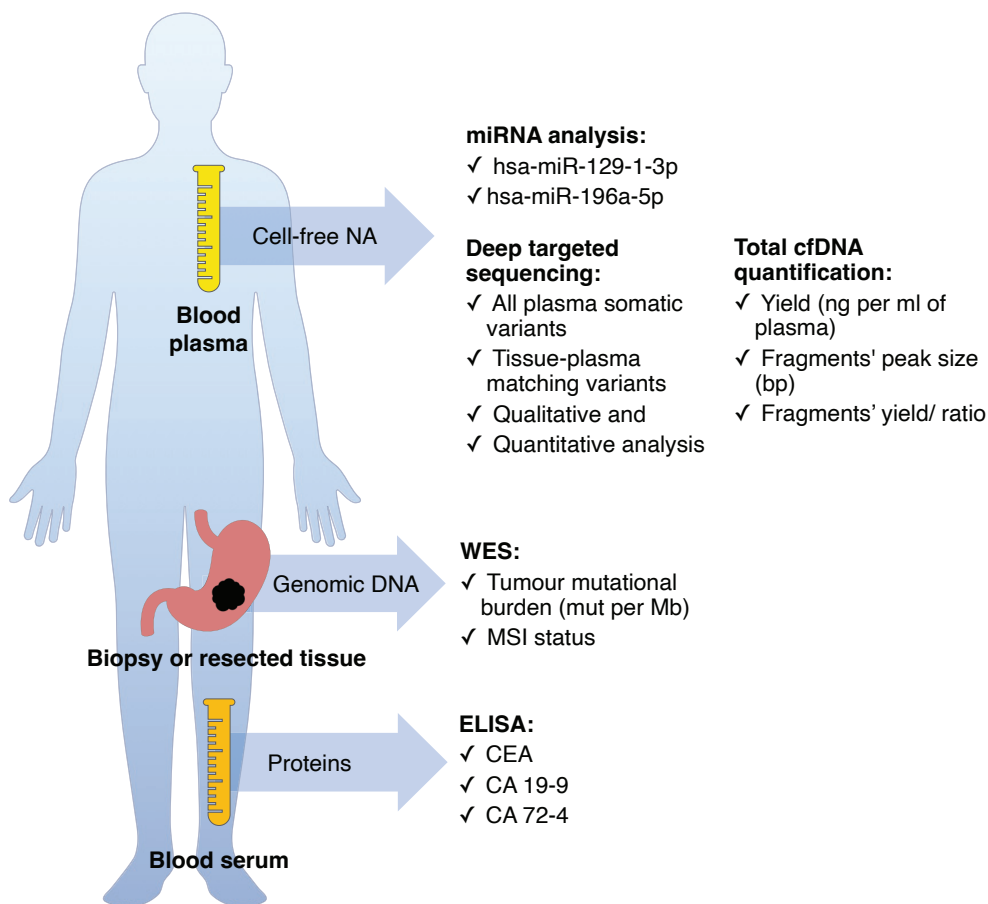
Fig. 2.5. *Multi-layer molecular analytes involved in the analysis.*

Survival analysis (n = 28) was performed including all of the described analytes. Data revealed that patients without somatic variants detected in cf-DNA had a median survival time (MST) of 803 days, patients with 1–2 cf-DNA sequence alterations had MST of 469 days, patients with 3–6 cfDNA sequence alterations – 315, and patients with more than 6–44 days (p = 0.008) (Fig. 2.6. A). To adjust for patients' demographics and tumours' characteristics such as age, gender, and size of the primary tumour based on tumour–node–metastasis staging, Cox proportional hazards model was used. Results showed a slight gender impact on survival estimation (padj = 0.0410) and a significant effect of more than 6 variants detected in plasma (padj = 0.0186) for a shorter lifespan. Moreover, hsa-miR-129-3p expression level showed a slight tendency of association with survival time (data not shown), and in combination with a quantity of tumour-derived cfDNA variants revealed that:

(1) patients without somatic variants and highly expressed hsa-miR-129-3p in plasma survived for an average of 894 days, and on the other hand (2) survival time of the patients with more than 6 tumour-derived cfDNA variants and low hsa-miR-129-3p expression in plasma was 44 days (p = 0.024 comparing all combinations and survival time) (Fig. 2.6. B).

Finally, the multilayer molecular profile (Fig. 2.5) was analysed for discrimination of the patients with larger tumours (T3-T4) and distant metastases. The analysis involved complete cases with full multi-layer molecular profiles (n = 17) and was performed based on feature selection and random forest classificator algorithms. Comparing T1-T1 *vs.* T3-T4 groups, the feature classificator showed that total cfDNA yield (ng/ml of plasma) was confirmed as an important attribute (maxImp = 10.20) while the concentration of CA 19-9 was assigned as a tentative attribute (maxImp = 5.17). In the case of M0 *vs.* M1 status, the feature classificator revealed that attributes such as expression level of hsa-miR-129-1-3p and quantity of overall detected alterations in plasma were confirmed as important (maxImp = 10.04, maxImp = 7.02 respectively). Qualitative detection of plasma cfDNA mutations was assigned as a tentative attribute (maxImp = 7.93) in this case.

Confirmed attributes were used for performance evaluation by using the receiver operating characteristic (ROC) curve. Total cfDNA yield resulted in an area under the curve (AUC) of 0.712 when discriminating between T1-T2 *vs.* T3-T4; on the other hand, the expression level of hsa-mir-129-1-3p and qualitative plasma cfDNA (presence of any somatic alteration) analysis could be implemented for discrimination between M0 *vs.* M1 (AUC = 0.818).
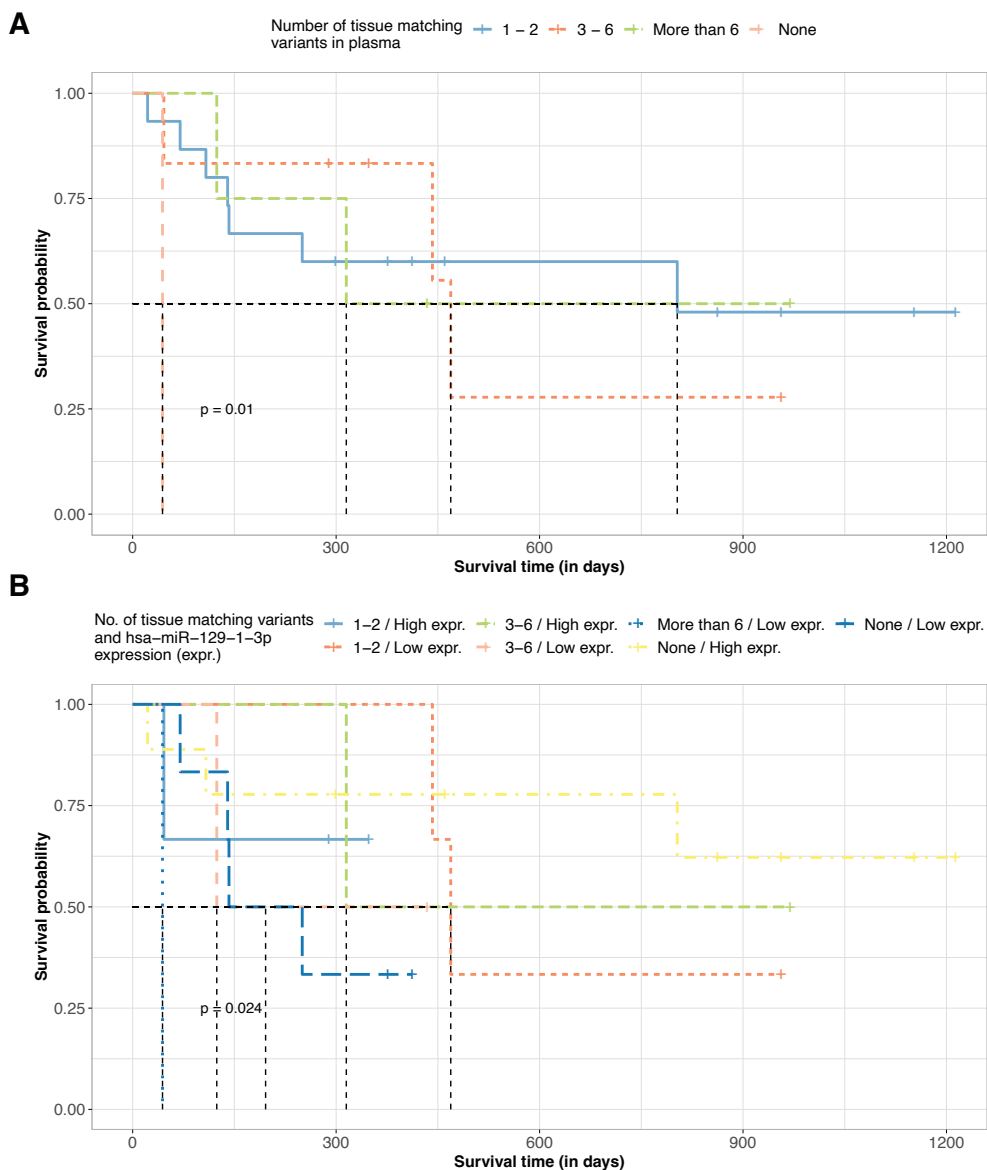
Fig. 2.6. *Kaplan-Meier survival analysis of GC patients with **A)** a different quantity of tumour-derived cfDNA alterations detected in plasma cfDNA: none (salmon), 1–2 (blue), 3–6 (orange), more than 6 (green); **B)** different combinations of high/ low expression of hsa-miR-129-1-3p and a quantity of tumour-derived cfDNA alterations detected in plasma cfDNA: no alterations and high expr. (yellow), no alterations and low expr. (dark blue), 1–2 and high expr. (light blue), 1-2 and low expr. (orange), 3–6 and high expr. (green), 3–6 and low expr. (salmon), more than 6 and low expr. (blue).*

3. DISCUSSION

The investigation of early diagnostic markers, especially minimally invasive, is an extremely relevant topic and in the future is expected to significantly improve GC patients' diagnostics, disease monitoring and consequently – survival. There is a wide variety of molecular candidates that could be detected and tracked during the course of a disease in different body fluids. Studies show that various miRNAs are associated with GC [46,64] or even its premalignant stages [45,65], and tissue miRNAs could be mirrored in the plasma and detected in a minimally invasive manner [66,67]. In addition to this, recent worldwide studies and even international projects such as GRAIL [68,69] started to focus on circulating tumour-derived DNA, which has a wide applicational spectra from the monitoring of a disease state [70]and relapse [71,72] to the prediction of a therapy effect [73]. All these findings lay a strong foundation for further cancer research and identification of minimally invasive potential target molecules or their combinations and also help to increase the fundamental understanding of underlying pathological molecular processes. Thus, this thesis presents a study consisting of two major parts analysing two types of plasma circulating molecules and in conjunction with computational biology directly or indirectly addresses the issues related to GC diagnosis and prognosis.

The first part of this thesis "**Epigenetic signatures**" has involved the following miRNA analysis aspects: (1) the evaluation of miRNA profiles of control, premalignant and malignant cases of gastric tissue, (2) hsa-miR-129-1-3p and hsa-miR-196a-5p expression evaluation in blood plasma, and (3) analysis of hsa-miR-20b-5p and hsa-miR-451a-5p role in gastric carcinogenesis.

To begin with, smRNA-seq data of control, premalignant, and malignant tissues has revealed both, repeatedly GC-related and novel miRNA biomarker targets. In concordance with other studies by Assumpcao *et al.* and Pereira *et al.* [74,75] our results have also confirmed miRNAs such as hsa-miR-3131, hsa-miR-483, hsa-miR-150, hsa-miR-200a-3p, hsa-miR-873-5p to be deregulated in GC. On the other hand, a number of previously not reported miRNAs such as hsa-miR-548ba, hsa-miR-4521, hsa-miR-549a have been identified. Although studies have suggested that hsa-miR-548ba could be associated with bladder cancer, hsa-miR-4521 – with *H. pylori* infection and hsa-miR-549a – with renal cancer [76–78], there is no data showing the potential association of these miRNAs with inflammatory or cancerous processes of gastric tissue. Analysis of AG miRNome has revealed deregulation of several miRNAs (*e. g.* hsa-miR-3591-3p, hsa-miR-122-3p/5p and hsa-miR-

451a), which have been previously described as biomarker candidates by Liu *et al.* [79]. However, the most deregulated miRNAs such as hsa-miR-215, hsa-miR-4497, hsa-miR-1251, *etc.* have been identified in AG tissue for the first time in our study. Previously miRNA hsa-miR-215-5p has been shown to be deregulated in different gastrointestinal tract pathologies such as Barrett's oesophagus, intraepithelial neoplastic lesions and ulcerative colitis [80–82], while hsa-miR-4497 and hsa-miR-452 have been associated with GC [83,84].

Second, we have shown that hsa-miR-129-1-3p and hsa-miR-196a-5p have been gradually deregulated in tissue samples corresponding to the CON-AG-GC sequence. Similarly to our results, Wang *et al.* and Yu *et al.* have also reported that hsa-miR-129-1-3p was down-regulated in GC tissues, have a tumour suppressing role and reflect the tissue expression pattern in the gastric juice [85,86]. Although hsa-miR-196a over-expression in AG tissue was not previously reported, other authors' results suggested that hsa-miR-196a is overexpressed in GC tissue, plasma, and commercial cell cultures [87,88]. We have, therefore, analysed the expression of hsa-miR-129-1-3p and hsa-miR-196a-5p in an independent cohort of CON, AG, and GC plasma samples, and revealed similar expression patterns compared to tissue samples. Significant hsa-miR-129-1-3p expression differences have been determined comparing AG and GC groups suggesting its potential as a minimally invasive diagnostic analyte.

Further analysis has involved the functional role of hsa-miR-20b-5p and hsa-miR-451a-5p, which, according to literature, have been shown to be deregulated in many different malignancies, including GC, and have a high potential to become novel therapeutic targets and/or biomarkers [52–58]. In concordance with the literature [55,56,89,90], our results have also revealed that hsa-miR-20b-5p is overexpressed in GC tissue samples compared to control. To our best knowledge, currently, there is no study analysing levels of hsa-miR-20b-5p in plasma samples. Similarly, our results of pilot smRNA-seq of plasma samples have shown no supporting sequencing reads of this miRNA suggesting that hsa-miR-20b-5p is not delivered to the circulation. On the other hand, studies have reported that miR-451a has been often down-regulated in a series of cancers including GC [91–93], [53,58] and this is also confirmed in our study. In addition, we were able to detect hsa-miR-451a-5p expression signals in plasma samples by the pilot smRNA-seq study (data not shown) which is consistent with studies that have been conducted by Lario *et al.* and Jiang *et al.* [59,60]. Finally, we have determined that hsa-miR-20b-5p and hsa-miR-451a-5p directly regulate the expression of genes possibly involved in the PI3K/AKT/mTOR pathway and loss-of-function experiments have revealed a significant tumour-suppressive role in GC.

To sum up, this study has shown that miRNA profiling could help to identify promising epigenetic biomarkers and plasma samples could mirror tissue miRNome expression patterns. On the other hand, miRNA functional studies are of high importance not only fundamentally but also for translational applications as it may help to improve the knowledge of molecular processes in carcinogenesis and the development of new treatment strategies.

For a more detailed discussion on the miRNA analysis results please refer to the papers related to the study of the Epigenetic signatures *("Atrophic gastritis and gastric cancer tissue miRNome analysis reveal hsa-miR-129-1 and hsa-miR-196a as potential early diagnostic biomarkers", "miR-20b and miR-451a Are Involved in Gastric Carcinogenesis through the PI3K/AKT/ mTOR Signaling Pathway: Data from Gastric Cancer Patients, Cell Lines and Ins-Gas Mouse Model")*.

The second section of the thesis "**Genetic signatures**" has involved extensive cfDNA analysis including the following aspects: (1) analysis of different biological materials for cfDNA isolation; (2) impact of preanalytical plasma processing variables; (3) approach for cfDNA sequencing; (4) plasma cfDNA mutational profile evaluation for GC patients.

After benchmarking of two commercial urine cfDNA isolation kits (NEXTprep-Mag Urine cfDNA Isolation Kit (PerkinElmer), Urine Cell-Free Circulating DNA Purification Midi Kit (Norgen Biotek)) we have revealed important specifications of each kit with different advantages and disadvantages considering cfDNA yield, processing time, price per sample, *etc*. Results have shown a vast variation in isolated DNA yields between both, different individuals, and separate days for the same individuals. We have also observed gender-related differences in cfDNA yields which are consistent with previous studies [94,95]. Strong signal of DNA fractions longer than known ~180 bp fragments have been detected suggesting that this fraction possibly derives from the epithelium of the urinary tract, after shedding and subsequent lysis of epithelial cells. Consequently, these longer fragments could cause wild-type contamination in the sample. For the above-mentioned reasons, using the urine cfDNA for the rare somatic analysis of cancers could be extremely challenging. Therefore, current recommendations suggest that plasma samples are a superior biological source for the cfDNA analysis [96]. In addition, plasma sampling is minimally invasive and well standardized.

The isolation procedure of plasma cfDNA could be affected by many pre-processing variables that should be considered. Thus, we have studied the impact of several important sample processing variables (*e. g.* isolation kit, storage, hemolysis, *etc.*) that could affect cfDNA and wild type gDNA yields. All three isolation kits tested (NextPrep-Mag cfDNA Isolation Kit (Bioo Scientific), MagMAX Cell-Free DNA Isolation Kit (Thermo Fisher Scienti-

fic), QIAamp Circulating Nucleic Acid Kit (Qiagen)) performed differently and should be selected according to the user's applications. Currently, QIAamp Isolation Kit is reported to be the most recommended for cfNA isolation [57,97–102], and we have also shown that a significantly greater amount of cfDNA was detected when quantified by the fluorescence-based method. Moreover, we have found that QIAamp Isolation Kit was more suitable for simultaneous circulating miRNA isolation in plasma.

Furthermore, we have compared different cfDNA sequencing approaches and analysed thresholds of artefacts by using cfDNA Horizon Reference Standard Set (Wild Type cfDNA Reference Standard and the 0.1 % cfDNA Reference Standard). Sequencing results of libraries with and without UMI have revealed that despite the loss of sequencing depth in UMI libraries, signal noise is significantly reduced. Results are concordant to the literature which shows improved accuracy of sequencing when using molecular tagging approach [61,62]. While low biomass samples require a lot of PCR cycles, the use of UMI libraries significantly reduce PCR errors. Also, read collapsing by UMI families allows to correctly quantify the AF and detect rare somatic alterations.

By implementing the above-described benchmarking and cfDNA analysis optimisation results, we have performed an analysis of mutational profiles of GC tissue and plasma samples. We have included the investigation of plasma cfDNA yield, fragmentation, correlation with oncoproteins and clinical data. Our cfDNA yield and fragmentation analysis have supported the previous findings by other studies [103–106], revealing a dramatic increase in total cfDNA yield, moderate positive correlation with serum CEA level, and smaller mono- and di-nucleosomal fragments for GC patients compared to the control. Although a high effort was made to accurately describe cancers in the molecular level (The Cancer Genome Atlas Research Network) [69], there is a lack of studies and gene panels for cfDNA analysis which could be applicable for everyday cancer diagnostics. Thus, we have evaluated the somatic mutational profile for GC tissue (WES sequencing) and developed a unique 38-gene panel. The panel consisted of GC related genes that were also reported in the COSMIC database [107] and was used for plasma cfDNA sample sequencing with a UMI tagging approach. Studies that were conducted by other authors have shown that the fraction of cancer patients with detectable tumour-derived somatic mutations in plasma cfDNA varies between 33.9 % to 58 % [41,108,109], and in comparison, our data have shown the concordance rate of 47.8 %. Overall, somatic variants in plasma cfDNA have been detected for the solid proportion of the patients: 21 of 23 alteration-positive tissue samples (91.3 %). In addition to this, quantitative analysis of tumour-derived alterations tracked in plasma showed that these variants have been

detected more often for patients with more advanced tumours (T3-T4 stage). Similar findings were reported by Hamakawa *et al.* and Fang *et al.* [41,43]. However, it is important to note, that even a relatively small (T1-T2) GC tumour could shed detectable ctDNA signals and in our case consisted of 10 % of the study cohort.

For a more detailed discussion on the results please refer to the papers related to the study of Genetic signatures (*"Quantifying cell free DNA in urine: comparison between commercial kits, impact of gender and inter-individual variation", "Effects of Quantification Methods, Isolation Kits, Plasma Biobanking, and Hemolysis on Cell-Free DNA Analysis in Plasma", "Quantifying sequencing error and effective sequencing depth of liquid biopsy NGS with UMI error correction",* and *"Liquid biopsy in gastric cancer: analysis of somatic cancer tissue mutations in plasma cell-free DNA for predicting disease state and patient survival"*).

Finally, in the third part "**Multi-layer** molecular analysis" we have summarised data from "Epigenetic signatures" and "Genetic signatures". Complete GC patients' cases were used for survival and discriminant analysis (T1-T2 *vs.* T3-T4 and M0 *vs.* M1). Survival analysis has revealed significant results when stratifying the groups by a quantity of tumour-derived somatic variants detected in plasma cfDNA. Also, a significant association with worse survival was revealed when levels of hsa-miR-129-1-3p were combined with quantitative analysis of GC-derived somatic variants in plasma samples. To date, worse overall survival for GC patients with increased total cfDNA yield [110], low cfDNA methylation level [111] and detection of ctDNA [112] was shown. In the terms of circulating miRNAs, although altered levels of several plasma miRNAs (*e. g.* miR-21, miR-23b [113]) have already been linked with the GC prognosis before, miRNA hsa-miR-129-1-3p alone or in combination with other analytes have been associated with GC patients' survival for the first time in our study.

In addition, by implementing the computational biology methods, we have estimated the importance of multi-level molecular analytes for discriminative analysis. We have shown that the expression level of hsa-mir-129-1-3p and the detected somatic alterations in the plasma cfDNA could potentially be applied for discrimination between patients with and without distant metastasis (AUC = 0.818). In comparison, Bu *et al.* [110] have shown that plasma cfDNA yield performed well in the prediction of distant metastases (AUC = 0.756), while in our analysis it was shown to possibly be an important attribute for discrimination between T1-T2 *vs.* T3-T4. Recent efforts for improving GC patients' care have included the multi-omics approach and have tried to address the issues related to response to neoadjuvant chemotherapy [49] or early diagnostics [50]. However, to our best knowledge, this is

the first study that has performed a comprehensive multi-layer analyte (epigenetic and genetic) analysis for GC prognosis and disease state.

Despite promising results that were presented in our study, there are some limitations that need to be addressed. First, plasma cfDNA analysis in samples of control and premalignant stages would help to estimate the sensitivity and specificity of custom gene panel for GC diagnostics. Also, the study sample size is small, and the study of Genetic signatures was not validated by different approach such as droplet digital PCR or in independent cohort of samples. However, it is important to note that the study population was well clinically defined and tested for many multi-layer clinically relevant variables. Moreover, a study involving an independent study population (GC patients' plasma samples collected during the course of a disease) is carried out at the moment by our group. Despite the above-mentioned limitations, we believe that this study adds very important new data for the development of clinically relevant liquid biopsy tools in patients with GC.

To conclude, circulating plasma miRNA and cfDNA remains one of the most frequently investigated molecular analytes. Our results have shown that by implementing sensitive sequencing-based analysis, molecular tools such as UMI tagging and multi-layer molecular modelling approaches, specific epigenetic and genetic alterations could be detected in the plasma of GC patients, helping to predict survival and disease state. Plasma sampling is a convenient and well-standardized procedure suitable for liquid biopsy in many different cancers and, although plasma cfNA analysis for routine applications is still challenging, analysis of multi-layer molecular analytes shows a great potential for minimally invasive tumour genome and/or epigenome characterisation, diagnosis, and prognosis.

SUMMARY OF CONCLUSIONS

1. Gastric tissue small RNA sequencing revealed 129 (82 up-regulated and 47 down-regulated) differentially expressed miRNAs when comparing gastric cancer and controls, 99 (67 up-regulated and 32 down-regulated) – gastric cancer and atrophic gastritis, and 20 (6 up-regulated and 14 down-regulated) - atrophic gastritis and controls.

2.1. Gastric tissue small RNA sequencing data showed that hsa-miR-129-1-3p and hsa-miR-196a-5p were deregulated gradually: following control, atrophic gastritis, and gastric cancer pathological cascade. Only hsa-miR-129-1-3p was significantly down-regulated in blood plasma samples when comparing atrophic gastritis and gastric cancer groups.

2.2. Expressions of hsa-miR-20b-5p and hsa-miR-451a-5p were significantly deregulated in gastric cancer tissue compared to controls. Experimental down-regulation of hsa-miR-20b-5p and up-regulation of hsa-miR-451a-5p *in vitro* revealed tumour-suppressive role by possibly targeting genes involved in PI3K/AKT/mTOR signalling pathway.

3. Whole-exome sequencing and very deep targeted sequencing data in tissue and plasma samples, respectively, revealed a 47.8 % concordance rate of the mutational profiles. Although, tumour-derived cell-free DNA sequence alterations were detected more often in samples of the patients with larger tumours (T3-T4), mutations were also detected for patients with T1-T2 status.

4. The increasing quantity of tumour-derived mutations in cell-free DNA and low expression level of hsa-miR-129-1-3p were associated with the worse patients' survival. Low expression level of hsa-mir-129-1-3p and the presence of somatic alterations in plasma could be potentially applied for discrimination of patients with distant metastasis (AUC = 0.818).

PRACTICAL RECOMMENDATIONS

In this work we have conducted several benchmark studies that address the cfDNA isolation related challenges and, therefore, enable us to indicate important practical recommendations that could be useful in both laboratory-research and translational-clinical, fields.

1. When designing a study, we recommend large sample amounts of urine to be collected. For NGS with 50 ng of cfDNA, we recommend collecting 60–70 mL urine.

2. Different fragment yields of urine cfDNA is isolated by using different commercially available isolation kits. It is critical to estimate fragment distribution by automated capillary electrophoresis (high-sensitivity reagents) and reconsider samples with strong signal of a peak > 1000 bp, which possibly appear from lysed normal epithelial cell of urinary tract. Moreover, researchers must be aware that this peak indicates potential source of gDNA contamination, which could complicate detection of rare somatic variants.

3. Above-mentioned points suggest that urine cfDNA analysis is still challenging and more reliable results could be obtained by using standardized collection of blood plasma. We recommend collection of plasma into tubes with K_2EDTA stabilizer (10 mL, Cat. No. 366643), and process with centrifugation within 1 hour. Double centrifugation and smooth brake profile should be used to prevent disruption of the buffy coat layer and possible gDNA contamination (first centrifugation step: room temperature at $2000 \times g$ for 10 minutes with smooth brake profile; second: $3000 \times g$ for 10 minutes with smooth brake profile).

4. We recommend plasma cfDNA quantification by automated capillary electrophoresis (high-sensitivity reagents) selecting the size range between 100 bp and 400 bp manually to estimate cfDNA yield. Recommended cfDNA amount depends on application, however 40-50 ng per ml of plasma on average for cancer patients could be normally obtained and one blood collection tube is enough. However, cfDNA yield in healthy individuals is extremely low and this should be considered when designing a study (two or more blood collection tubes could be necessary). Reconsider inclusion of samples with prominent peak > 1000 bp as described in paragraph no. 1.

5. UMI tagging approach requires very deep target sequencing (Illumina recommends sequencing to 35,000× minimum raw sequencing depth), however, this approach allows confident rare variant calling and quantification (in case VAF ~0.3 %). Paired cancer or normal tissue samples would help to confidently discriminate between sequencing artifacts and/or true somatic variants.

SANTRAUKA

Įvadas. Laisvai cirkuliuojančios nukleorūgštys (lcNR) į kraujotaką patenka ląstelės apoptozės, nekrozės ar net aktyvios sekrecijos metu ir gali atspindėti ląstelės, iš kurios yra kilusi, genetiką ir epigenetiką [8]–[11]. Tyrimai, kuriuose dalyvavo sveiki asmenys, rodo, kad didžiausią lcNR frakciją sudaro laisvai cirkuliuojanti DNR (lcDNR), kilusi iš kraujodaros ląstelių [10]. LcDNR fragmento dydis yra apie 180 bazių porų (bp) [12]–[14] ir atitinka DNR apsivijusios aplink nukleosomą ilgį [15]. Tačiau tyrimai rodo, kad galima ir kitokia lcDNR fragmentacija: trumpesni fragmentai aptinkami auglio, vaisiaus kilmės lcDNR [13], [16] ir organų transplantacijos atveju [17]. Kai lcDNR patenka į kraujotaką, jos šalinimas yra ypač greitas: nuo 16 minučių iki 2,5 valandos [18]–[20]. Svarbu pažymėti, kad ši lcDNR savybė laikoma viena iš svarbiausių, ypač šio molekulinio įrankio pritaikymui stebėti ligos eigą „realiu laiku" ar kuomet reikalingas pakartotinis ir ilgalaikis mėginių ėmimas pacientams.

Nuolatinis ligos būklės stebėjimas yra labai svarbus vėžiu sergančių pacientų medicininei priežiūrai. Taigi, navikinės kilmės lcDNR – dar vadinama cirkuliuojančia naviko DNR (cnDNR) – gali būti labai perspektyvus molekulinis įrankis ir padėti sumažinti invazinių intervencijų poreikį. CnDNR kraujotakoje gali atsirasti iš pirminių arba metastazavusių navikų bei cirkuliuojančių naviko ląstelių. Remiantis literatūra, cnDNR dalis kraujotakoje svyruoja nuo mažiau nei 0,1 % iki daugiau nei 25 % [21], [22]. Yra parodyta, kad lcDNR kiekis yra susijęs su naviko išplitimu, jo atsinaujinimu ir paciento atsaku į gydymą [21], [23], [24]. Naujausi technologiniai sprendimai leidžia atlikti išsamesnę cnDNR analizę. Naujos kartos sekvenavimas (NKS) įgalina taikyti taikininę specifinių genų sekoskaitą (angl. *targeted gene sequencing*) ar net analizuoti visą genomo seką (angl. *whole genome sequencing*). Tai leidžia nustatyti daugybę pakitimų, įskaitant chromosominius pokyčius, genų amplifikacijas ir kitus genų persitvarkymus [6], [25]–[28].

Cirkuliuojančios DNR mutacijos yra genetiniai pokyčiai, atspindintys tik dalį su naviku susijusių molekulinių procesų. Kraujo mikroRNR (miRNR) profiliavimas taip pat gali suteikti labai vertingos epigenetinės informacijos. Šiandien yra žinoma daugiau nei 2500 miRNR, kurios yra siejamos su svarbiais biologiniais procesais (2021 m. gruodžio mėn. „miRBase" versija 22) [29]. MiRNR yra puikus kandidatas tapti plačiai taikomu biožymeniu: miRNR yra viena gausiausių RNR molekulių kraujyje, be to, mažas miRNR molekulių dydis lemia jų stabilumą ir leidžia atlikti patikimą analizę daugelyje skirtingų biologinių skysčių [30]. Atlikti tyrimai rodo, kad kraujyje cirkuliuojančių miRNR profilis atspindi naviko audinio profilį, o tai gali būti

pritaikoma minimaliai invazinei diagnostikai [31], [32]. Svarbus ne tik diagnostinis, bet ir terapinis miRNR potencialas, dėl šios priežasties miRNR funkciniai tyrimai, apimantys genų taikinių nustatymą ir *in vitro* analizę, yra ypatingos svarbos.

Skrandžio vėžys (SV) yra viena iš labiausiai paplitusių virškinamojo trakto onkologinių ligų Lietuvoje ir visame pasaulyje [34]. Jis dažnai diagnozuojamas jau pažengusiose stadijose, o per 2020-uosius sąlygojo 768 tūkst. žmonių mirčių visame pasaulyje (2020 m. GLOBOCAN duomenys) [34]. Daugeliu atvejų SV vystymasis yra laipsniškas procesas ir gali apimti tokias būkles kaip lėtinis skrandžio gleivinės uždegimas, atrofinis gastritas (AG), žarnyno metaplazija (ŽM) ir galiausiai SV. Tai kompleksinė liga, atsirandanti dėl aplinkos (pvz., mityba, didelis druskos suvartojimas ir kt.) ir paties organizmo (pvz., *H. Pylori* infekcija, genetika, amžius ir kt.) veiksnių sąveikos.

Tik nedidelė dalis (apie 10 %) SV atvejų turi teigiamą šeiminę anamnezę ir tik 1–3 % turi nustatomas lytinių ląstelių mutacijas – daugiausia *CDH1* gene, kuris koduoja E-kadherino baltymą [35]. Daugeliu atvejų paveldimą ligą lemiantys genetiniai pakitimai lieka nežinomi. Per pastarąjį dešimtmetį atlikti tyrimai prisidėjo prie išsamesnio SV molekulinio profilio nustatymo ir parodė, kad procesą inicijuojantys genetiniai pokyčiai dažniausiai aptinkami onkogenuose, tokiuose kaip *KRAS*, *PIK3CA*, *CNTBB1*, *ERBB3* ir kt., ir naviko supresoriuose, tokiuose kaip *TP53*, *CDH1*, *ARID1A*, *FAT4* ir kt. [38].

Nepaisant instrumentinių diagnostinių tyrimų (endoskopinė biopsija ir kompiuterinė tomografija, pozitronų emisijos tomografija ir kt.) įvairovės, vis dar yra keletas labai svarbių iššūkių SV diagnostikoje: mirtingumas, susijęs su ligos diagnozavimu pažengusioje stadijoje, bei sunkumai nustatant komplikacijas, tokias kaip minimali liekamoji liga ar naviko išplitimas į pilvaplėvės ertmę [39]. Įprastų diagnostikos metodų bei šiuo metu naudojamų molekulinių biožymenų vaidmuo, diagnozuojant SV ankstyvose stadijose ir nustatant komplikacijas, išlieka labai ribotas. Be to, standartinės biopsijos susijusios su įvairiomis komplikacijomis [40]. Apibendrinant, šiuo metu vis dar labai reikalingi minimaliai invaziniai biožymenys, kurie padėtų pagerinti SV pacientų medicininę priežiūrą, šios ligos diagnostiką ir stebėjimą. Daugelis tyrimų rodo, kad taikant genetinį [41]–[43] ir epigenetinį [44]–[46] molekulinį profiliavimą, skystosios biopsijos tyrimai galėtų reikšmingai prisidėti sprendžiant daugelį su SV diagnostika ir prognoze susijusių problemų. Tyrimai taip pat rodo svarbią pažangą taikant daugiasluoksnę molekulinę analizę diagnostikai. Pavyzdžiui, Kohenas su bendraautoriais sukūrė baltymų ir lcDNR vertinimu pagrįstą *CancerSEEK* aštuonių dažnai sutinkamų navikų nustatymo testą, kurio jautrumas siekia nuo 69 % iki 98 %, o specifiškumas > 99 % [50]. Tačiau SV atveju nėra tyrimų, įtraukiančių epigenetinį, genetinį ir baltymų lygmenis į daugiasluoksnę molekulinę analizę.

Šiame tyrime vertinome skystosios biopsijos analizės pritaikomumą skrandžio navikų tyrimams. Šiam tikslui pasirinkome išsamų įvairius lygmenis apimantį studijos dizainą: nuo tinkamiausios biologinės medžiagos ir kraujo paruošimo įtakos iki plazmos ir audinių molekulinių profilių palyginimo ir net funkcinės molekulių analizės *in vitro*. Pilnam SV ir AG audinių miRNR profiliavimui, pakitusios raiškos miRNR validavimui plazmoje ir plazmos lcDNR mutacijų profilio nustatymui naudojome aukšto našumo ir tikslumo molekulinius metodus (pvz.: mažųjų RNR sekoskaitą, viso egzomo sekoskaitą, realaus laiko PGR (RL-PGR) ir kt.). Visa tai leido įgyvendinti šį tyrimą ir pateikti optimizuotus kraujo apdorojimo protokolus, nustatyti miRNR, kurios pakinta gastrito-karcinomos eigos metu, parodyti dviejų plačiai tirtų miRNR vaidmenį skrandžio kancerogenezėje, įvertinti plazmos lcDNR somatinius pakitimus ir atskleisti, kad daugiasluoksnė molekulinė analizė gali būti perspektyvus SV eigos stebėjimo ir prognozės įrankis.

Darbo tikslas. Šiuo tyrimu buvo siekiama analizuoti laisvai cirkuliuojančias nukleorūgštis skrandžio vėžiu sergančių pacientų kraujo plazmoje ir įvertinti šių biožymenų funkciją bei tinkamumą minimaliai invazinei diagnostikai ir prognozei.

Darbo uždaviniai:
1. Ištirti pacientų, sergančių atrofiniu gastritu ir skrandžio vėžiu, mikroRNR profilį audiniuose, lyginant su kontroline grupe.
2. Validuoti labiausiai pakitusių su audiniu susijusių mikroRNR raišką skrandžio vėžio pacientų kraujo plazmoje ir įvertinti specifinių mikroRNR funkcinį vaidmenį skrandžio kancerogenezėje.
3. Nustatyti skrandžio audinio DNR ir plazmos laisvai cirkuliuojančios DNR mutacijų profilius skrandžio vėžiu sergantiems pacientams ir įvertinti plazmos laisvai cirkuliuojančios DNR sąsają su klinikiniais duomenimis.
4. Įvertinti daugiasluoksnio molekulinio profiliavimo pritaikomumą skrandžio vėžio prognozei ir ligos būklės stebėjimui.

Mokslinis darbo naujumas ir aktualumas. Šiame tyrime pateikiami: (1) skrandžio ikinavikinių (AG), navikinių (SV) ir kontrolinių (KON) audinių diferenciniai miRNR raiškos profiliai bei labiausiai dereguliuotų miRNR analizė kraujo plazmoje; (2) naujos įžvalgos, atskleidžiančios svarbų hsa-miR-20b-5p ir hsa-miR-451a-5p vaidmenį skrandžio kancerogenezėje, ir galinčios pasitarnauti ateityje kuriant skrandžio navikų gydymo strategiją; (3) įžvalgos apie šlapimo surinkimą skystosios biopsijos tyrimui ir lcDNR analizei; (4) optimizuoti lcNR išskyrimo protokolai ir išsamus ėminio laikymo ir ap-

dorojimo sąlygų (saugojimas, hemolizė, išskyrimo metodas), galinčių turėti įtakos lcNR kokybei, aprašymas; (5) dviejų plazmos lcDNR sekos nustatymo metodų palyginimas, siekiant patikimai aptikti žemo dažnio somatinius variantus; (6) SV audinių ir plazmos mutacijų profilių persidengimo ir galimo pritaikomumo ligos stebėjimui ir išgyvenamumui analizė.

Apibendrinant, tai yra pirmasis tyrimas, kuriame buvo pateikti išsamūs europietiškos kilmės tiriamųjų ikinavikinių ir navikinių skrandžio audinių miR-NR profiliai, parodytas funkcinis specifinių epigenetinių biožymenų (hsa-miR-20b-5p) vaidmuo skrandžio kancerogenezėje, įvertintas SV plazmos ir audinių mutacijų profilių persidengimas, naudojant unikalius molekulinius žymenis (UMŽ) (angl. *unique molecular identifiers* (UMI)) ir gilią sekoskaitą, bei SV eigos statuso ir išgyvenamumo analizėje įdiegta 14 daugiasluoksnių molekulinių analičių, apjungiančių keletą molekulinių lygmenų: genetinį, epigenetinį ir baltymų. Šis išsamus tyrimas pateikia svarbių rezultatų, susijusių su ėminio apdorojimu ir sekoskaitos metodo parinkimu, atsižvelgiant į skystosios biopsijos pritaikymą SV diagnozei ir prognozei.

Autoriaus indėlis. Žemiau pagal su disertacija susijusias publikacijas (A1–A6, pateiktas skyriuje *List of scientific papers*) pateikiamas autorės Gretos Varkalaitės indėlis.

A1: prisidėjo prie studijos koncepcijos ir tyrimo planavimo, tyrimo dalyvių klinikinių ir fenotipinių duomenų rinkimo, atliko visus tyrimus ir laboratorinius eksperimentus (bendros RNR ir lcDNR išskyrimas, kokybės kontrolė, RL-PGR), duomenų analizę (mažųjų RNR sekoskaitos duomenų analizė, diferencinė miRNR raiškos ir statistinė analizė), ir parengė publikaciją. Naudoti bioinformatiniai įrankiai ir paketai plačiau aprašyti skyriuje *1.5 Next-generation sequencing and bioinformatics data analysis section.*

A2: atlikti laboratoriniai eksperimentai (RL-PGR, baltymų tyrimas *Western Blot* metodu, ląstelių kultūrų liposominė transfekcija, MTT, kolonijų formavimo, apoptozės, žaizdos gijimo (angl. *Wound healing)* tyrimai), duomenų analizė (*in silico* miRNR genų taikinių numatymas, statistinė analizė) ir duomenų vizualizacija, parengė pirminį publikacijos rankraštį, prisidėjo prie straipsnio rašymo, recenzavimo ir redagavimo proceso.

A3: atliko lcDNR išskyrimą (šlapimo mėginiai), kokybės kontrolę ir kiekybinį lcDNR vertinimą, lcDNR sekoskaitos bibliotekų konstravimą ir sekoskaitos eksperimentus, išanalizavo duomenis ir prisidėjo prie publikacijos rengimo.

A4: atliko visus laboratorinius eksperimentus ir tyrimus (kraujo plazmos paruošimas, lcDNR išskyrimas iš plazmos mėginių, kokybės kontro-

lė ir kiekybinis vertinimas, RL-PGR), statistię duomenų analizę ir vizualizaciją, parengė publikaciją.

A5: buvo atsakinga už duomenų analizę (lcDNR UMŽ sekoskaitos duomenų analizė), prisidėjo prie publikacijos rengimo ir gavo pilną atsakingos autorės (angl. *corresponding author*) sutikimą naudoti ir apginti šį straipsnį savo disertacijoje. Naudoti bioinformatiniai įrankiai ir paketai plačiau aprašyti skyriuje *1.5 Next-generation sequencing and bioinformatics data analysis section*.

A6: atliko laboratorinius eksperimentus (DNR išskyrimas iš audinių mėginių, lcDNR išskyrimas iš plazmos mėginių, ELISA, bibliotekų paruošimas viso egzomo sekoskaitai ir tikslinei lcDNR sekoskaitai), klinikinių ir fenotipinių tyrimo dalyvių duomenų rinkimą, duomenų analizę ir interpretavimą (viso egzomo ir lcDNR UMŽ sekoskaitos duomenų analizė, genų sąrašo signalinių kelių praturtinimo, diskriminacinė analizė, pagrįsta daugiakomponenčiu analičių modeliavimu, statistinė analizė), vizualizavo duomenis, paruošė publikaciją. Naudoti bioinformatiniai įrankiai ir paketai plačiau aprašyti skyriuje *1.5 Next-generation sequencing and bioinformatics data analysis section*.

Medžiagos ir metodai. Disertacijoje atliktiems tyrimams gauti atitinkamų institucijų etikos leidimai (Nr. BE-2-10, Nr. BE-2-31 ir Nr. B327/10, Nr. D470/14) (1-2 ir 3-4 priedai). Tiriamoji populiacija detaliai aprašyta atitinkamose publikacijose: "*Atrophic gastritis and gastric cancer tissue miRNome analysis reveal hsa-miR-129-1 and hsa-miR-196a as potential early diagnostic biomarkers*", "*miR-20b and miR-451a Are Involved in Gastric Carcinogenesis through the PI3K/AKT/mTOR Signaling Pathway: Data from Gastric Cancer Patients, Cell Lines and Ins-Gas Mouse Model*", "*Quantifying cell free DNA in urine: comparison between commercial kits, impact of gender and inter-individual variation*" and "*Effects of Quantification Methods, Isolation Kits, Plasma Biobanking, and Hemolysis on Cell-Free DNA Analysis in Plasma*", "*Liquid biopsy in gastric cancer: analysis of somatic cancer tissue mutations in plasma cell-free DNA for predicting disease state and patient survival*". Genominė DNR (gDNR) iš SV audinių ir pacientų kraujo leukocitų bei visuminė RNR buvo išskirta naudojant komercinius rinkinius: *AllPrep DNA/RNA Mini Kit* (*Qiagen*) ir išdruskinimo (angl. *salting-out*) metodą, bei *miRNeasy Mini Kit* arba *miRNeasy Micro Kit (Qiagen)*. Šlapimo lcDNR skyrimui naudoti du komerciniai rinkiniai: (1) *NEXTprep-Mag Urine cfDNA Isolation Kit (PerkinElmer)*; (2) *Urine Cell-Free Circulating DNA Purification Midi Kit (Norgen Biotek)*. Taip pat, plazmos lcDNR: (1) *NextPrep-Mag cf-*

DNA izoliavimo rinkinys (Bioo Scientific); (2) *MagMAX Cell-Free DNA Isolation Kit (Thermo Fisher Scientific)*; (3) *QIAamp Circulating Nucleic Acid Kit (Qiagen)*. Visos procedūros buvo atliekamos pagal gamintojo instrukcijas.

Kiekybinė RL-PGR buvo naudojami šie *TaqMan* mikroRNR zondai ir pradmenys: hsa-miR-129* (ID: 002298), hsa-miR-196a (ID: 241070_mat), has-miR-223 (ID: 002098), hsa-miR-20b 5p (ID: 001014) ir hsa-miR-451a-5p (ID: 001141), RL-PGR atlikta naudojant 7500 *Fast* realaus laiko PGR sistemą (*Applied Biosystems*). Slenkstinio ciklo (Ct) reikšmės buvo normalizuotos naudojant RNU6B (ID: 0010930) arba has-miR-16 (ID: 000391) (*Thermo Fisher Scientific*) endogenines kontroles.

Mažųjų RNR sekoskaitos bibliotekos buvo paruoštos naudojant *Illumina TruSeq Small RNA Sample Preparation Kit (Illumina)* ir sekvenuotos naudojant *Illumina HiSeq 2500* platformą pagal gamintojo protokolą. Mažųjų RNR sekoskaitos duomenų analizė atlikta naudojant nf-core/smrnaseq analizės algoritmą v.1.0.0. SV audinių ir to paties paciento kraujo leukocitų mėginių egzomo sekoskaita buvo atlikta naudojant *Illumina TruSeq Nano DNA Library Prep Kit (Illumina)* ir *Integrated DNA Technologies xGen Exome Research Panel (Integrated DNA Technologies)*. Labai gilus sekos nustatymas (40 000 ×) buvo atliktas plazmos lcDNR, o bibliotekos sukonstruotos naudojant *TruSight Oncology UMI* reagentus (*Illumina*) ir *Integrated DNA Technologies xGen Custom Panel (Integrated DNA Technologies)*. Mėginiai buvo sekvenuoti naudojant *NovaSeq 6000* platformą pagal gamintojo instrukcijas. *GATK Best Practices paired-sample* [51] analizės algoritmas buvo naudojama egzomo sekos duomenų analizei, o plazmos lcDNR sekos analizė buvo atlikta naudojant *Illumina UMI Error Correction App* (v1.0.0.1) bioinformatinį įrankį.

AGS ir MKN28 ląstelių linijų transfekcijai ir funkcinei analizei naudotos sintetinės miRNR: hsa-miR-451a-5p, hsa-miR-20b-5, anti-miR-20b ir miRNR neigiama kontrolė (*miRVana, Ambion by Thermo Fisher Scientific*). Potencialių miRNR genų taikinių analizė atlikta šia tvarka: (1) *in silico* analizė (*DIANA Lab Tools TarBase, miRanda, TargetScan*), (2) baltymų pokyčių analizė *Western Blot* metodu ir (3) luciferazės reporterio tyrimas. Funkciniai ląstelių tyrimai apėmė MTT, kolonijų formavimo, tėkmės citometrijos (*Annexin V/ PI*) bei žaizdų gijimo tyrimus. *Western Blot* tyrime buvo naudojami šie antikūnai: IRF1 (ab186384; *Abcam*), PTEN (ab32199, *Abcam*), TXNIP (40-3700; Thermo Fisher Scientific), CAV1 (ab192869, *Abcam*), TSC1 (37-0400, *Thermo Fisher Scientific*), ir GAPDH (AM4300; *Ambion by Thermo Fisher Scientific*). Serumo onkobaltymams nustatyti buvo naudojamas su fermentais susietas imunosorbentinis tyrimas (ELISA) ir šie komerciniai rinkiniai: žmogaus karcinoembrioninio antigeno (CEA) (ab99992; *Abcam*), žmogaus vėžio

antigeno CA 19-9 (ab108642; *Abcam*) ir žmogaus vėžio antigeno CA 72-4 (E-EL-H0613; *Elabscience*).

Rezultatai ir jų apžvalga. Ankstyvųjų diagnostinių žymenų, ypač minimaliai invazinių, tyrimas yra itin aktuali tema dėl kurios ateityje tikimasi žymiai pagerinti SV pacientų diagnostiką, ligos eigos stebėjimą ir išgyvenamumą. Tyrimai rodo, jog yra daug įvairių molekulių, kurias galima aptikti ir sekti ligos eigoje skirtinguose kūno skysčiuose. Studijų rezultatai atskleidė, kad įvairios miRNR yra susijusios su SV [46], [64] ar net ikinavikinėmis SV būklėmis [45], [65], o audinių miRNR gali būti atspindimos kraujo plazmoje ir aptiktos minimaliai invaziniu būdu [66], [66], [67]. Naujausiuose pasauliniuose tyrimuose ir net tarptautiniuose projektuose, tokiuose kaip *GRA-IL* [68], [69], daugiausia dėmesio skiriama cnDNR, kuri turi platų taikymo spektrą nuo ligos nustatymo [70], jos atsinaujinimo [71], [72] iki terapijos efekto [73] vertinimo. Visi šie tyrimai sudaro tvirtą pagrindą tolesnei analizei ir minimaliai invazinių molekulių ar jų derinių identifikavimui vėžiu sergantiems pacientams, tai taip pat padeda geriau suprasti pagrindinius patologinius molekulinius procesus, vykstančius kancerogenezės metu. Taigi, šioje disertacijoje pristatomas tyrimas, susidedantis iš dviejų pagrindinių dalių, analizuojančių dviejų tipų cirkuliuojančias molekules plazmoje ir įtraukiančių bioinformatinių įrankių pritaikymą, tiesiogiai arba netiesiogiai sprendžia su SV diagnoze ir prognoze susijusias problemas.

Pirmoji šio darbo dalis „Epigenetiniai žymenys" apėmė šiuos miRNR analizės aspektus: (1) kontrolinių, ikinavikinių ir navikinių skrandžio audinio grupių miRNR profilių įvertinimą, (2) hsa-miR-129-1-3p ir hsa- miR-196a-5p raiškos įvertinimą kraujo plazmoje ir (3) hsa-miR-20b-5p ir hsa-miR-451a-5p vaidmens skrandžio kancerogenezėje analizę.

Pirmiausia, kontrolinės grupės (KON), ikinavikinių (atrofinio gastrito (AG)) ir navikinių būklių skrandžio audinių mažųjų RNR sekoskaitos duomenys atskleidė tiek anksčiau su SV sietus, tiek ir naujus miRNR biožymenis-kandidatus. Mūsų rezultatai sutampa su Assumpcao ir bendraautorių bei Pereira ir bendraautorių [74], [75] tyrimais ir taip pat patvirtina, kad miRNR, tokios kaip hsa-miR-3131, hsa-miR-483, hsa-miR-150, hsa-miR-200a-3p, hsa-miR-873-5p įprastai yra dereguliuotos SV atveju. Kita vertus, buvo nustatyta ir keletas anksčiau su SV nesietų miRNR, tokių kaip hsa-miR-548ba, hsa-miR-4521, hsa-miR-549a. Nors tyrimai rodo, kad hsa-miR-548ba gali būti susijusi su šlapimo pūslės vėžiu, hsa-miR-4521 – su *H. Pylori* infekcija, o hsa-miR-549a – su inkstų vėžiu [76]–[78], duomenų, rodančių galimą šių miRNR ryšį su uždegiminiais ar navikiniais skrandžio audinio pakitimais šiuo metu nėra. AG miRNR profiliavimo analizė parodė šių miRNR dere-

guliaciją: pvz., hsa-miR-3591-3p, hsa-miR-122-3p/5p ir hsa-miR-451a, kurios anksčiau jau buvo aprašytos Liu ir bendraautorių tyrime kaip galimi SV biožymenys [79]. Tačiau AG audinyje labiausiai pakitusios raiškos miRNR, tokios kaip hsa-miR-215, hsa-miR-4497, hsa-miR-1251 ir kt., mūsų tyrime buvo nustatytos pirmą kartą. Kitų tyrėjų buvo parodyta, kad miRNR hsa-miR-215-5p yra dereguliuota sergant kitomis virškinimo trakto patologijomis, tokiomis kaip Barreto stemplė, intraepiteliniai neoplastiniai pažeidimai ir opinis kolitas [80]–[82], o hsa-miR-4497 ir hsa-452 anksčiau buvo susietos tik su SV [83], [84].

Antra, mes parodėme, kad hsa-miR-129-1-3p ir hsa-miR-196a-5p raiškos lygis yra palaipsniui pakitęs audinių mėginiuose ir atitinka KON-AG-SV patogenezės seką. Panašiai kaip ir mūsų tyrime, Wang ir bendraautoriai bei Yu su kolegomis taip pat parodė, kad hsa-miR-129-1-3p raiška buvo sumažėjusi SV audiniuose, taip pat ši miRNR pasižymi naviką slopinančiomis savybėmis, o jos ekspresijos lygis audinyje atsispindi ir skrandžio sultyse [85], [86]. Nors anksčiau nebuvo atlikta tyrimų rodančių padidėjusią hsa-miR-196a raišką AG audinyje, kitų autorių rezultatai atskleidžia, kad hsa-miR-196a raiška gali būti padidėjusi navikiniame skrandžio audinyje, SV pacientų plazmoje ir komercinėse SV ląstelių kultūrose [87], [88]. Dėl šios priežasties mes analizavome hsa-miR-129-1-3p ir hsa-miR-196a-5p ekspresiją nepriklausomoje KON, AG ir SV plazmos mėginių grupėje ir parodėme panašų šių miRNR raiškos lygį, lyginant su audinių mėginiais. Palyginus AG ir SV pacientų grupes, buvo nustatyti statistiškai reikšmingi hsa-miR-129-1-3p raiškos skirtumai, o tai rodo šios molekulės, kaip minimaliai invazinės diagnostinės analitės, potencialą.

Tolesnė analizė buvo susijusi su hsa-miR-20b-5p ir hsa-miR-451a-5p funkcinio vaidmens tyrimu. Remiantis literatūra, šios miRNR yra dereguliuotos daugeliu skirtingų piktybinių navikų atveju, įskaitant SV, ir turi didelį potencialą tapti naujais terapijos taikiniais ir (arba) navikiniais biožymenimis [52]–[58]. Kaip ir kiti tyrėjai [55], [56], [89], [90], mes taip pat parodėme, kad hsa-miR-20b-5p raiška yra padidėjusi SV audinių mėginiuose, palyginti su kontroline grupe. Mūsų žiniomis, šiuo metu nėra tyrimų, kurie būtų vertinę hsa-miR-20b-5p lygį kraujo plazmos mėginiuose. Atitinkamai, mūsų pilotinės studijos, įtraukusios plazmos mėginių mažųjų RNA sekoskaitą, rezultatai neparodė šios miRNR raišką patvirtinančių nuskaitymų, o tai leidžia manyti, kad hsa-miR-20b-5p nepatenka į kraujotaką. Kita vertus, tyrimai rodo, kad hsa-miR-451a-5p raiška daugelio navikų atveju, įskaitant ir SV, yra sumažėjusi [91]–[93], [53], [58], ir tai taip pat patvirtina mūsų gauti rezultatai. Be to, mums pavyko nustatyti hsa-miR-451a-5p raiškos signalus plazmos mėginiuose, atliekant pilotinę mažųjų RNR sekoskaitos studiją (duomenys nepateikti). Tai atitinka rezultatus ir kitų studijų, kurias atliko Lario ir bendraautoriai bei Jiang su kolegomis [59], [60]. Galiausiai nustatėme, kad hsa-

miR-20b-5p ir hsa-miR-451a-5p tiesiogiai reguliuoja genų, galimai dalyvaujančių PI3K/AKT/mTOR signaliniame kelyje, raišką, o funkcijos praradimo (angl. *loss-of-function*) *in vitro* eksperimentai atskleidė reikšmingą navikus slopinantį poveikį SV atveju.

Apibendrinant, šis tyrimas parodė, kad miRNR profiliavimas gali padėti nustatyti potencialius epigenetinius biožymenis, o plazmos mėginiai gali atspindėti audinių miRNR raiškos pokyčius. Kita vertus, miRNR funkciniai tyrimai yra labai svarbūs ne tik fundamentaliai, bet ir vertinant tyrimų pritaikomumą: jie gali padėti pagerinti žinias apie vykstančius molekulinius procesus ir padėti kurti naujas terapijos strategijas.

Išsamesnės diskusijos apie miRNR analizės rezultatus ieškokite publikuotuose straipsniuose, susijusiuose su epigenetinių žymenų tyrimu *("Atrophic gastritis and gastric cancer tissue miRNome analysis reveal hsa-miR-129-1 and hsa-miR-196a as potential early diagnostic biomarkers", "miR-20b and miR-451a Are Involved in Gastric Carcinogenesis through the PI3K/AKT/mTOR Signaling Pathway: Data from Gastric Cancer Patients, Cell Lines and Ins-Gas Mouse Model"*).

Antroje disertacijos dalyje „Genetiniai žymenys" buvo atlikta išsami lcDNR analizė, apimanti šiuos aspektus: (1) skirtingų biologinių medžiagų tinkamų lcDNR išskyrimui analizę; (2) preanalitinių plazmos mėginių veiksnių vertinimą; (3) lcDNR sekoskaitos metodo pasirinkimą; (4) plazmos lcDNR mutacijų profilio įvertinimą pacientams, sergantiems SV.

Atlikę dviejų komercinių šlapimo lcDNR išskyrimo rinkinių (*NEXT-prep-Mag Urine cfDNA Isolation Kit (PerkinElmer), Urine Cell-Free Circulating DNA Purification Midi Kit (Norgen Biotek)*) palyginimą, atskleidėme svarbias kiekvieno rinkinio specifikacijas, skirtingus privalumus ir trūkumus, atsižvelgdami į bendrą lcDNR kiekį, mėginio apdorojimo laiką, vieno mėginio kainą ir t.t. Rezultatai parodė žymius lcDNR kiekio skirtumus tarp skirtingų individų ir skirtingų tyrimo atlikimo dienų net tiems patiems asmenims. Taip pat pastebėjome su lytimi susijusius lcDNR kiekio skirtumus, kurie yra aprašyti ir ankščiau atliktuose tyrimuose [94], [95]. Buvo aptiktas stiprus DNR frakcijų, ilgesnių nei žinomi ~ 180 bp fragmentai, signalas, leidžiantis manyti, kad ši frakcija gali kilti iš šlapimo takų epitelio po epitelio ląstelių lizės. Šie ilgesni fragmentai gali sukelti užteršimą laukinio tipo genomine DNR (gDNR) mėginyje. Dėl minėtų priežasčių šlapimo lcDNR naudojimas retų somatinių mutacijų analizei gali būti labai sudėtingas. Todėl dabartinėse rekomendacijose teigiama, kad plazmos mėginiai yra patikimesnis biologinis šaltinis lcDNR analizei [96]. Be to, plazmos mėginių ėmimas yra minimaliai invazinis ir gerai standartizuotas.

Plazmos lcDNR išskyrimo procedūrai įtakos gali turėti daugelis preanalitinių veiksnių, į kuriuos reikėtų atsižvelgti. Taigi, mes ištyrėme kelių svarbių

kintamųjų, tokių kaip išskyrimo rinkinio, mėginio saugojimo, hemolizės ir kt., galinčių turėti įtakos lcDNR ir laukinio tipo gDNR kiekiui, poveikį. Visi trys išbandyti išskyrimo rinkiniai (*NextPrep-Mag cfDNA Isolation Kit (Bioo Scientific), MagMAX Cell-Free DNA Isolation Kit (Thermo Fisher Scientific), QIAamp Circulating Nucleic Acid Kit (Qiagen)*) pasižymėjo skirtingomis savybėmis ir turėtų būti pasirenkami pagal poreikį. Kitų autorių tyrimai rodo, kad QIAamp išskyrimo rinkinys yra dažniausiai rekomenduojamas lcNR išgryninimui [57], [97]–[102]. Mes parodėme, kad šiuo rinkiniu išskiriamas reikšmingai didesnis lcDNR kiekis vertinant koncentraciją išmatuotą fluorescenciniu metodu. Be to, buvo parodyta, kad QIAamp išskyrimo rinkinys yra tinkamesnis kuomet tuo pačiu metu izoliuojama cirkuliuojanti miRNR plazmoje.

Taip pat, palyginome skirtingus lcDNR sekoskaitos metodus ir išanalizavome sekoskaitos artefaktų dažnį, o šiam tikslui naudojome *cfDNA Horizon Reference Standard Set* rinkinį, kuri sudaro laukinio tipo lcDNR referentinis standartas ir 0,1 % lcDNR referentinis standartas. Sekoskaitos bibliotekų su ir be UMŽ sekomis rezultatai atskleidė, kad nepaisant sekos gylio praradimo UMŽ bibliotekose, signalo triukšmas yra reikšmingai sumažinamas. Šie rezultatai atitinka ir literatūros duomenis, kur parodytas daug geresnis sekoskaitos nustatymo tikslumas, naudojant UMŽ molekulinio žymėjimo metodą [61], [62]. Nors ruošiant sekoskaitos bibliotekas mažos biomasės mėginiams reikalinga daugiau PGR ciklų, UMŽ naudojimas žymiai sumažina PGR klaidas. Be to, analizės metu po UMŽ šeimų perdengimo, galima tiksliau įvertinti alelių dažnį ir aptikti retas somatines mutacijas.

Įdiegę aukščiau aprašytus lcDNR analizės optimizavimo rezultatus, atlikome SV audinių ir plazmos mėginių mutacijų profilių analizę. Į tyrimą įtraukėme plazmos lcDNR kiekį, fragmentaciją, koreliacijos su onkobaltymais ir klinikiniais duomenimis analizę. Bendro lcDNR kiekio ir fragmentacijos vertinimas patvirtino ankstesnius kitų tyrėjų rezultatus [103]–[106] ir atskleidė ryškų bendro lcDNR kiekio padidėjimą, vidutinę teigiamą koreliaciją su serumo CEA baltymo lygiu ir mažesnius mono- ir di-nukleosominius fragmentus SV pacientams, palyginti su kontrole. Nors pastaruoju metu dedamos didžiulės tyrėjų pastangos kuo tiksliau apibūdinti molekulinį vėžio profilį (*The Cancer Genome Atlas Research Network*) [69], vis dar trūksta tyrimų ir sudarytų genų rinkinių (panelių), skirtų lcDNR analizei, kurios galėtų būti pritaikomos kasdieninėje vėžio diagnostikoje. Todėl įvertinome SV audinio somatinių mutacijų profilį (pagal viso egzomo sekoskaitos duomenis) ir sukūrėme unikalų 38 genų rinkinį. Rinkinį sudarė su SV susiję genai, aprašyti *COSMIC* (angl. *Catalogue of Somatic Mutations in Cancer*) duomenų bazėje [107]. Šis genų rinkinys buvo naudojamas plazmos lcDNR mėginių sekoskaitos metu kartu naudojant UMŽ žymėjimo metodą bei gilią sekoskaitą. Kitų

autorių atlikti tyrimai rodo, kad vėžiu sergančių pacientų pirminio naviko kilmės somatinės mutacijos plazmos lcDNR dalis svyruoja nuo 33,9 % iki 58 % [41], [108], [109], o tuo tarpu mūsų duomenys parodė 47,8 % dažnį. Apskritai, somatiniai variantai plazmos lcDNR buvo aptikti nemažai pacientų daliai: 21 iš 23 pacientų su mutacijomis nustatytomis audinyje (91,3 %). Be to, kiekybinė navikinės kilmės mutacijų plazmoje analizė parodė, kad šie variantai dažniau aptinkami pacientams, turintiems labiau pažengusius navikus (T3-T4 stadija). Panašius rezultatus savo studijoje parodė ir Hamakawa su bendraautoriais bei Fang su tyrėjų grupe [41], [43]. Tačiau svarbu pažymėti, kad mūsų studijoje taip pat buvo nustatyta, kad net santykinai maži SV navikai gali produkuoti aptinkamą cnDNR kiekį (10 % tiriamųjų).

Išsamesnę diskusiją apie rezultatus rasite publikacijose, susijusiose su genetinių žymenų tyrimu (*"Quantifying cell free DNA in urine: comparison between commercial kits, impact of gender and inter-individual variation"*, *"Effects of Quantification Methods, Isolation Kits, Plasma Biobanking, and Hemolysis on Cell-Free DNA Analysis in Plasma"*, *"Quantifying sequencing error and effective sequencing depth of liquid biopsy NGS with UMI error correction"*, and *"Liquid biopsy in gastric cancer: analysis of somatic cancer tissue mutations in plasma cell-free DNA for predicting disease state and patient survival"*).

Galiausiai surinkome ir apibendrinome dviejų pagrindinių tyrimo dalių duomenis. Išgyvenamumo ir ROC (angl. *receiver operating characteristic curve*) kreivės (T1-T2 *vs.* T3-T4 ir M0 *vs.* M1) analizei buvo naudojami pilnai charakterizuoti SV pacientų atvejai. Išgyvenamumo analizė atskleidė reikšmingus rezultatus stratifikuojant SV pacientų grupes pagal navikinės kilmės mutacijų aptiktų plazmoje kiekį. Be to, reikšmingas ryšys su blogesniu išgyvenimu buvo parodytas, kartu vertinant hsa-miR-129-1-3p raiškos lygį ir kiekybinę navikinės kilmės mutacijų plazmoje analizę. Iki šiol kitų tyrėjų nustatytas prastesnis bendras SV pacientų išgyvenamumas buvo stebimas padidėjus bendram lcDNR kiekiui [110], esant žemam lcDNR metilinimo lygiui [111] ir kuomet nustatoma cnDNR [112]. Kalbant apie cirkuliuojančias miR-NR, nors pakitęs kelių plazmos miRNR lygis (pvz., miR-21, miR-23b [113]) tyrimuose siejamas su SV prognoze, miRNR hsa-miR-129-1-3p ekspresija (atskirai arba kartu su kitomis analitėmis) mūsų tyrime su SV pacientų išgyvenimu susieta pirmą kartą.

Be to, taikydami bioinformatinius modelius, įvertinome daugiasluoksnės molekulinės analizės svarbą pasitelkiant ROC kreives. Buvo parodyta, kad hsa-mir-129-1-3p raiškos lygis ir nustatytos somatinės plazmos lcDNR mutacijos gali būti naudojamos siekiant atskirti pacientus, turinčius ir neturinčius tolimų metastazių (plotas po ROC kreive (AUC) = 0,818). Bu ir bendraautorių [110] studija parodė, kad bendras plazmos lcDNR kiekis geru tikslumu

gali padėti nustatyti pacientus su tolimomis metastazėmis (AUC = 0,756), tuo tarpu mūsų analizė atskleidė, kad tai gali būti svarbus žymuo atskiriant T1-T2 ir T3-T4 stadijos pacientus. Pastarųjų metų tyrėjų pastangos gerinti SV pacientų medicininę priežiūrą įtraukia daugiasluoknę molekulinę analizę, bandant spręsti problemas, susijusias su atsako į neoadjuvantinę chemoterapiją vertinimu [49] arba ankstyva diagnostika [50]. Tačiau, mūsų žiniomis, tai yra pirmasis tyrimas, kurio metu atlikta išsami daugiasluoksnė molekulinė (epigenetinė ir genetinė) analizė vertinant SV prognozę ir ligos būklę.

Nepaisant daug žadančių rezultatų, kurie buvo pateikti mūsų tyrime, yra keletas apribojimų, į kuriuos reikia atsižvelgti. Pirma, plazmos lcDNR analizė kontrolinių ir ikinavikinių stadijų mėginiuose padėtų įvertinti mūsų sudaryto genų rinkinio sekoskaitos jautrumą ir specifiškumą SV diagnostikai. Be to, tyrimo imties dydis yra mažas, o genetinių žymenų tyrimas nebuvo validuotas skirtingu metodu (pvz., skaitmeninė lašelių PGR) arba nepriklausomoje mėginių grupėje. Nepaisant to, svarbu pažymėti, kad tyrimo populiacija buvo gerai kliniškai apibūdinta, atlikta daugiasluoksnė molekulinė kliniškai reikšmingų analičių analizė. Be to, šiuo metu mūsų grupė atlieka tyrimą, kuriame dalyvauja nepriklausoma tyrimo populiacija (SV pacientų plazmos mėginiai, paimti ligos eigoje). Nepaisant aukščiau paminėtų apribojimų, manome, kad šis tyrimas prideda labai svarbių naujų duomenų, skirtų kliniškai reikšmingų skystosios biopsijos įrankių kūrimui pacientams, sergantiems SV.

Apibendrinant galima pasakyti, kad cirkuliuojančios plazmos miRNR ir lcDNR išlieka vienos iš dažniausiai tiriamų molekulinių analičių. Mūsų rezultatai parodė, kad pritaikant jautrią naujos kartos sekoskaitos technologiją, tokius molekulinius įrankius kaip UMI žymėjimas bei daugiasluoksnius molekulinio modeliavimo metodus, SV pacientų plazmoje galima aptikti specifinius epigenetinius ir genetinius pakitimus, kurie gali padėti įvertinti pacientų išgyvenamumą ir ligos stadiją. Plazmos mėginių ėmimas yra patogi ir gerai standartizuota procedūra, tinkama daugelio skirtingų vėžio tipų skystosios biopsijos aplikacijoms, ir nors plazmos lcNR analizės taikymas kaip rutininio tyrimo vis dar kelia nemažai iššūkių, daugiasluoksnė molekulinių analičių analizė rodo didelį potencialą kaip minimaliai invazinis naviko genomo ir (arba) epigenomo vertinimo, diagnozės ir prognozės nustatymo įrankis.

Išvados

1. Palyginus skrandžio audinio mažų RNR sekoskaitos duomenis tarp skrandžio vėžio ir kontrolinių asmenų nustatytos 129 (82 padidėjusios ir 47 sumažėjusios), tarp skrandžio vėžio ir atrofinio gastrito – 99 (67 padidėjusios ir 32 sumažėjusios), tarp atrofinio gastrito ir kontrolinės grupės – 20 (6 padidėjusios ir 14 sumažėjusios) pakitusios raiškos miRNR.

2.1. Skrandžio audinio mažųjų RNR sekoskaitos duomenys taip pat parodė, kad hsa-miR-129-1-3p ir hsa-miR-196a-5p raiška buvo laipsniškai pakitusi ir atitiko kontrolės, atrofinio gastrito ir skrandžio vėžio patologinę seką. Tik hsa-miR-129-1-3p raiška buvo reikšmingai sumažėjusi lyginant atrofinio gastrito ir skrandžio vėžio kraujo plazmos mėginius.

2.2. Hsa-miR-20b-5p ir hsa-miR-451a-5p raiška buvo statistiškai reikšmingai pakitusi skrandžio vėžio audinyje, palyginus su kontroliniu audiniu. *In vitro* tyrimuose eksperimentiškai padidinus hsa-miR-20b-5p kiekį ir sumažinus hsa-miR-451a-5p kiekį nustatytas naviką slopinantis šių miRNR poveikis, galimai per genus, dalyvaujančius PI3K/AKT/mTOR signaliniame kelyje.

3. Audinio mėginių viso egzomo sekoskaitos ir kraujo plazmos mėginių labai gilios taikininės sekoskaitos duomenys, atskleidė 47,8 % mutacijų profilių persidengimą. Nors somatiniai laisvai cirkuliuojančios DNR sekos pakitimai (atitinkantys navikinio audinio mutacijas) dažniau nustatyti didesnius navikus (T3-T4) turinčių pacientų plazmos mėginiuose, somatinės plazmos DNR mutacijos buvo aptiktos ir pacientams su T1-T2 stadijos navikais.

4. Didėjantis naviko kilmės mutacijų skaičius laisvai cirkuliuojančioje DNR ir žemas hsa-miR-129-1-3p raiškos lygis buvo susiję su prastesniu pacientų išgyvenamumu. Sumažėjusi hsa-mir-129-1-3p raiška ir somatinių mutacijų nustatymas kraujo plazmoje gali būti pritaikomas atskiriant pacientus, turinčius tolimųjų metastazių (AUC = 0,818).

REFERENCES

1. Mandel P and Métais P. Les acides nucléiques du plasma sanguin chez l'homme. C R Seances Soc Biol Fil 1948; 142:241–243.
2. Sidransky D, Von Eschenbach A, Tsai YC, Jones P, Summerhayes I, Marshall F, *et al.* Identification of p53 gene mutations in bladder cancers and urine samples. Science (80-) 1991; 252(5006):706–709.
3. Dennis Lo YM, Corbetta N, Chamberlain PF, Rai V, Sargent IL, Redman CWG and Wainscoat JS. Presence of fetal DNA in maternal plasma and serum. Lancet 1997; 350(9076):485–487.
4. De Vlaminck I, Valantine HA, Snyder TM, Strehl C, Cohen G, Luikart H, *et al.* Circulating cell-free DNA enables noninvasive diagnosis of heart transplant rejection. Sci Transl Med 2014; 6(241).
5. De Vlaminck I, Martin L, Kertesz M, Patel K, Kowarsky M, Strehl C, *et al.* Noninvasive monitoring of infection and rejection after lung transplantation. Proc Natl Acad Sci U S A 2015; 112(43):13336–13341.
6. Wan JCM, Massie C, Garcia-Corbacho J, Mouliere F, Brenton JD, Caldas C, *et al.* Liquid biopsies come of age: Towards implementation of circulating tumour DNA. Nat Rev Cancer 2017; 17(4):223–238.
7. Alix-Panabières C, Schwarzenbach H and Pantel K. Circulating tumor cells and circulating tumor DNA. Annu Rev Med 2012; 63:199–215.
8. Thierry AR, El Messaoudi S, Gahan PB, Anker P and Stroun M. Origins, structures, and functions of circulating DNA in oncology. Cancer Metastasis Rev 2016; 35(3):347–76.
9. Jahr S, Hentze H, Englisch S, Hardt D, Fackelmayer FO, Hesch R and Knippers R. DNA Fragments in the Blood Plasma of Cancer Patients : Quantitations and Evidence for Their Origin from Apoptotic and Necrotic Cells DNA Fragments in the Blood Plasma of Cancer Patients : Quantitations and Evidence for Their Origin from Apoptotic and Necr. Cancer Res 2001; 61(4):1659–1665.
10. Lehmann-Werman R, Neiman D, Zemmour H, Moss J, Magenheim J, Vaknin-Dembinsky A, *et al.* Identification of tissue-specific cell death using methylation patterns of circulating DNA. Proc Natl Acad Sci 2016; 113(13):E1826–E1834.
11. Snyder MW, Kircher M, Hill AJ, Daza RM and Shendure J. Cell-free DNA Comprises an In Vivo Nucleosome Footprint that Informs Its Tissues-Of-Origin. Cell 2016; 164(1–2):57–68.
12. Giacona MB, Ruben GC, Iczkowski KA, Roos TB, Porter DM and Sorenson GD. Cell-free DNA in human blood plasma: length measurements in patients with pancreatic cancer and healthy controls. Pancreas 1998; 17(1):89–97.
13. Lo YMD, Chan KCA, Sun H, Chen EZ, Jiang P, Lun FMF, *et al.* Maternal plasma DNA sequencing reveals the genome-wide genetic and mutational profile of the fetus. Sci Transl Med 2010; 2(61).
14. Thierry AR, Mouliere F, Gongora C, Ollier J, Robert B, Ychou M, *et al.* Origin and quantification of circulating DNA in mice with human colorectal cancer xenografts. Nucleic Acids Res 2010; 38(18):6159–6175.
15. Retureau R, Foloppe N, Elbahnsi A, Oguey C and Hartmann B. A dynamic view of DNA structure within the nucleosome: Biological implications. J Struct Biol 2020; 211(1):107511.
16. Underhill HR, Kitzman JO, Hellwig S, Welker NC, Daza R, Baker DN, *et al.* Fragment Length of Circulating Tumor DNA. PLoS Genet 2016; 12(7).

17. Zheng YWL, Chan KCA, Sun H, Jiang P, Su X, Chen EZ, *et al.* Nonhematopoietically derived DNA is shorter than hematopoietically derived DNA in plasma: a transplantation model. Clin Chem 2012; 58(3):549–558.

18. Shao JY, Li YH, Gao HY, Wu QL, Cui NJ, Zhang L, *et al.* Comparison of Plasma Epstein-Barr Virus (EBV) DNA Levels and Serum EBV Immunoglobulin A/Virus Capsid Antigen Antibody Titers in Patients with Nasopharyngeal Carcinoma. Cancer 2004; 100(6):1162–1170.

19. Dennis Lo YM, Zhang J, Leung TN, Lau TK, Chang AMZ and Magnus Hjelm N. Rapid clearance of fetal DNA from maternal plasma. Am J Hum Genet 1999; 64(1):218–224.

20. Yao W, Mei C, Nan X and Hui L. Evaluation and comparison of in vitro degradation kinetics of DNA in serum, urine and saliva: A qualitative study. Gene 2016; 590(1):142–148.

21. Diehl F, Schmidt K, Choti MA, Romans K, Goodman S, Li M, *et al.* Circulating mutant DNA to assess tumor dynamics. Nat Med 2008; 14(9):985–990.

22. Mouliere F, Robert B, Peyrotte E, Del Rio M, Ychou M, Molina F, *et al.* High fragmentation characterizes tumour-derived circulating DNA. PLoS One 2011; 6(9).

23. Leary RJ, Sausen M, Kinde I, Papadopoulos N, Carpten JD, Craig D, *et al.* Detection of chromosomal alterations in the circulation of cancer patients with whole-genome sequencing. Sci Transl Med 2012; 4(162).

24. Sorenson GD, Pribish DM, Valone FH, Memoli VA, Bzik DJ and Yao SL. Soluble normal and mutated DNA sequences from single-copy genes in human blood. Cancer Epidemiol Biomarkers Prev 3(1):67–71.

25. Forshew T, Murtaza M, Parkinson C, Gale D, Tsui DWY, Kaper F, *et al.* Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA. Sci Transl Med 2012; 4(136).

26. Leary RJ, Sausen M, Kinde I, Papadopoulos N, Carpten JD, Craig D, *et al.* Detection of chromosomal alterations in the circulation of cancer patients with whole-genome sequencing. Sci Transl Med 2012; 4(162).

27. Chan KCA, Jiang P, Zheng YWL, Liao GJW, Sun H, Wong J, *et al.* Cancer genome scanning in plasma: detection of tumor-associated copy number aberrations, single-nucleotide variants, and tumoral heterogeneity by massively parallel sequencing. Clin Chem 2013; 59(1):211–224.

28. Siravegna G, Marsoni S, Siena S and Bardelli A. Integrating liquid biopsies into the management of cancer. Nat Rev Clin Oncol 2017; 14(9):531–548.

29. Kozomara A, Birgaoanu M and Griffiths-Jones S. miRBase: from microRNA sequences to function. Nucleic Acids Res 2019; 47(D1):D155–D162.

30. Iorio M V. and Croce CM. MicroRNA dysregulation in cancer: diagnostics, monitoring and therapeutics. A comprehensive review. EMBO Mol Med 2012; 4(3):143–159.

31. Lawrie CH, Gal S, Dunlop HM, Pushkaran B, Liggins AP, Pulford K, *et al.* Detection of elevated levels of tumour-associated microRNAs in serum of patients with diffuse large B-cell lymphoma. Br J Haematol 2008; 141(5):672–675.

32. Ghafouri-Fard S, Vafaee R, Shoorei H and Taheri M. MicroRNAs in gastric cancer: Biomarkers and therapeutic targets. Gene 2020; 757:144937.

33. Cui C and Cui Q. The relationship of human tissue microRNAs with those from body fluids. Sci Rep 2020; 10(1).

34. Smyth EC, Nilsson M, Grabsch HI, van Grieken NC and Lordick F. Gastric cancer. Lancet 2020; 396(10251):635–648.

35. Oliveira C, Pinheiro H, Figueiredo J, Seruca R and Carneiro F. Familial gastric cancer: genetic susceptibility, pathology, and implications for management. Lancet Oncol 2015; 16(2):e60–e70.

36. Donner I, Kiviluoto T, Ristimäki A, Aaltonen LA and Vahteristo P. Exome sequencing reveals three novel candidate predisposition genes for diffuse gastric cancer. Fam Cancer 2015; 14(2):241–246.

37. Fewings E, Larionov A, Redman J, Goldgraben MA, Scarth J, Richardson S, *et al.* Germline pathogenic variants in PALB2 and other cancer-predisposing genes in families with hereditary diffuse gastric cancer without CDH1 mutation: a whole-exome sequencing study. lancet Gastroenterol Hepatol 2018; 3(7):489–498.

38. Tan P and Yeoh KG. Genetics and Molecular Pathogenesis of Gastric Adenocarcinoma. Gastroenterology 2015; 149(5):1153-1162.e3.

39. Yarema R, Ohorchak M, Hyrya P, Kovalchuk Y, Safiyan V, Karelin I, *et al.* Gastric cancer with peritoneal metastases: Efficiency of standard treatment methods. World J Gastrointest Oncol 2020; 12(5):569.

40. Yao MD, Rosenvinge EC von, Groden C and Mannon PJ. Multiple endoscopic biopsies in research subjects: safety results from a National Institutes of Health series. Gastrointest Endosc 2009; 69(4):906.

41. Fang W-L, Lan Y-T, Huang K-H, Liu C-A, Hung Y-P, Lin C-H, *et al.* Clinical significance of circulating plasma DNA in gastric cancer. Int J Cancer 2016; 138(12):2974–2983.

42. Wang DS, Liu ZX, Lu YX, Bao H, Wu X, Zeng ZL, *et al.* Liquid biopsies to track trastuzumab resistance in metastatic HER2-positive gastric cancer. Gut 2019; 68(7):1152–1161.

43. Hamakawa T, Kukita Y, Kurokawa Y, Miyazaki Y, Takahashi T, Yamasaki M, *et al.* Monitoring gastric cancer progression with circulating tumour DNA. Br J Cancer 2015; 112(2):352–356.

44. Komatsu S, Kiuchi J, Imamura T, Ichikawa D and Otsuji E. Circulating microRNAs as a liquid biopsy: a next-generation clinical biomarker for diagnosis of gastric cancer. J Cancer Metastasis Treat 2018; 4(7):36.

45. Zhu XL, Ren LF, Zhang L, Ding FH, Li X, Wang HP, *et al.* Plasma microRNAs as potential new biomarkers for early detection of early gastric cancer. World J Gastroenterol 2019; 25(13):1580–1591.

46. Link A and Kupcinskas J. MicroRNAs as non-invasive diagnostic biomarkers for gastric cancer: Current insights and future perspectives. World J Gastroenterol 2018; 24(30):3313–3329.

47. Nguyen QH, Nguyen H, Nguyen T and Le DH. Multi-Omics Analysis Detects Novel Prognostic Subgroups of Breast Cancer. Front Genet 2020; 11:1265.

48. Lindskrog SV, Prip F, Lamy P, Taber A, Groeneveld CS, Birkenkamp-Demtröder K, *et al.* An integrated multi-omics analysis identifies prognostic molecular subtypes of non-muscle-invasive bladder cancer. Nat Commun 2021 121 2021; 12(1):1–18.

49. Li Z, Gao X, Peng X, Chen MJM, Li Z, Wei B, *et al.* Multi-omics characterization of molecular features of gastric cancer correlated with response to neoadjuvant chemotherapy. Sci Adv 2020; 6(9).

50. Cohen JD, Li L, Wang Y, Thoburn C, Afsari B, Danilova L, *et al.* Detection and localization of surgically resectable cancers with a multi-analyte blood test. Science (80-) 2018; 359(6378):926–930.

51. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, *et al.* The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010; 20(9):1297–1303.

52. Ventura A, Young AG, Winslow MM, Lintault L, Meissner A, Erkeland SJ, *et al.* Targeted Deletion Reveals Essential and Overlapping Functions of the miR-17~92 Family of miRNA Clusters. Cell 2008; 132(5):875–886.

53. Li C-Y, Liang G-Y, Yao W-Z, Sui J, Shen X, Zhang Y-Q, *et al.* Identification and functional characterization of microRNAs reveal a potential role in gastric cancer progression. Clin Transl Oncol 2017; 19(2):162–172.

54. Pan X, Wang R and Wang Z-X. The Potential Role of miR-451 in Cancer Diagnosis, Prognosis, and Therapy. Mol Cancer Ther 2013; 12(7):1153–1162.

55. Xue T-M, Tao L, Zhang M, Xu G-C, Zhang J and Zhang P-J. miR-20b overexpression is predictive of poor prognosis in gastric cancer. Onco Targets Ther 2015; 8:1871–6.

56. Guo J, Miao Y, Xiao B, Huan R, Jiang Z, Meng D and Wang Y. Differential expression of microRNA species in human gastric cancer versus non-tumorous tissues. J Gastroenterol Hepatol 2009; 24(4):652–657.

57. Espinosa-Parrilla Y, Muñoz X, Bonet C, Garcia N, Venceslá A, Yiannakouris N, *et al.* Genetic association of gastric cancer with miRNA clusters including the cancer-related genes MIR29, MIR25, MIR93 and MIR106: results from the EPIC-EURGAST study. Int J cancer 2014; 135(9):2065–2076.

58. Riquelme I, Tapia O, Leal P, Sandoval A, Varga MG, Letelier P, *et al.* miR-101-2, miR-125b-2 and miR-451a act as potential tumor suppressors in gastric cancer through regulation of the PI3K/AKT/mTOR pathway. Cell Oncol 2016; 39(1):23–33.

59. Lario S, Brunet-Vega A, Quílez ME, Ramírez-Lázaro MJ, Lozano JJ, Garcia-Martínez L, *et al.* Expression profile of circulating microRNAs in the Correa pathway ofprogression to gastric cancer. United Eur Gastroenterol J 2018; 6(5):691.

60. Jiang X, Wang W, Yang Y, Du L, Yang X, Wang L, *et al.* Identification of circulating microRNA signatures as potential noninvasive biomarkers for prediction and prognosis of lymph node metastasis in gastric cancer. Oncotarget 2017; 8(39):65132–65142.

61. Kinde I, Wu J, Papadopoulos N, Kinzler KW and Vogelstein B. Detection and quantification of rare mutations with massively parallel sequencing. Proc Natl Acad Sci U S A 2011; 108(23):9530–9535.

62. Kivioja T, Vähärautio A, Karlsson K, Bonke M, Enge M, Linnarsson S and Taipale J. Counting absolute numbers of molecules using unique molecular identifiers. Nat Methods 2011; 9(1):72–74.

63. Thomas PD, Campbell MJ, Kejariwal A, Mi H, Karlak B, Daverman R, *et al.* PANTHER: A library of protein families and subfamilies indexed by function. Genome Res 2003; 13(9):2129–2141.

64. Juzenas S, Salteniene V, Kupcinskas J, Link A, Kiudelis G, Jonaitis L, *et al.* Analysis of deregulated micrornas and their target genes in gastric cancer. PLoS One 2015; 10(7).

65. Mao Y, Liu R, Zhou H, Yin S, Zhao Q, Ding X and Wang H. Transcriptome analysis of miRNA-lncRNA-mRNA interactions in the malignant transformation process of gastric cancer initiation. Cancer Gene Ther 2017; 24(6):267–275.

66. Liu J, Li Z, Teng W and Ye X. Identification of downregulated circRNAs from tissue and plasma of patients with gastric cancer and construction of a circRNA-miRNA-mRNA network. J Cell Biochem 2020; 121(11):4590–4600.

67. Zhang Y, Han T, Feng D, Li J, Wu M, Peng X, *et al.* Screening of non-invasive miRNA biomarker candidates for metastasis of gastric cancer by small RNA sequencing of plasma exosomes. Carcinogenesis 2020; 41(5):582–590.

68. Cohen JD, Javed AA, Thoburn C, Wong F, Tie J, Gibbs P, *et al.* Combined circulating tumor DNA and protein biomarker-based liquid biopsy for the earlier detection of pancreatic cancers. Proc Natl Acad Sci U S A 2017; 114(38):10202–10207.

69. Bass AJ, Thorsson V, Shmulevich I, Reynolds SM, Miller M, Bernard B, *et al.* Comprehensive molecular characterization of gastric adenocarcinoma. Nature 2014; 513(7517):202–209.

70. Abbosh C, Birkbak NJ, Wilson GA, Jamal-Hanjani M, Constantin T, Salari R, *et al.* Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. Nature 2017; 545(7655):446–451.

71. Tie J, Wang Y, Tomasetti C, Li L, Springer S, Kinde I, *et al.* Circulating tumor DNA analysis detects minimal residual disease and predicts recurrence in patients with stage II colon cancer. Sci Transl Med 2016; 8(346).

72. Azad TD, Chaudhuri AA, Fang P, Qiao Y, Esfahani MS, Chabon JJ, *et al.* Circulating Tumor DNA Analysis for Detection of Minimal Residual Disease After Chemoradiotherapy for Localized Esophageal Cancer. Gastroenterology 2020; 158(3):494-505.e6.

73. Magbanua MJM, Swigart LB, Wu HT, Hirst GL, Yau C, Wolf DM, *et al.* Circulating tumor DNA in neoadjuvant-treated breast cancer reflects response and survival. Ann Oncol Off J Eur Soc Med Oncol 2021; 32(2):229–239.

74. Pereira A, Moreira F, Vinasco-Sandoval T, Cunha A, Vidal A, Ribeiro-Dos-Santos AM, *et al.* miRNome Reveals New Insights Into the Molecular Biology of Field Cancerization in Gastric Cancer. Front Genet 2019; 10(JUN).

75. Assumpção MB, Moreira FC, Hamoy IG, Magalhães L, Vidal A, Pereira A, *et al.* High-Throughput miRNA Sequencing Reveals a Field Effect in Gastric Cancer and Suggests an Epigenetic Network Mechanism. Bioinform Biol Insights 2015; 9:111–117.

76. Zhao F, Ge YZ, Zhou LH, Xu LW, Xu Z, Ping WW, *et al.* Identification of hub miRNA biomarkers for bladder cancer by weighted gene coexpression network analysis. Onco Targets Ther 2017; 10:5551–5559.

77. Xuan Z, Chen C, Tang W, Ye S, Zheng J, Zhao Y, *et al.* TKI-Resistant Renal Cancer Secretes Low-Level Exosomal miR-549a to Induce Vascular Permeability and Angiogenesis to Promote Tumor Metastasis. Front cell Dev Biol 2021; 9.

78. Teng G, Dai Y, Chu Y, Li J, Zhang H, Wu T, *et al.* Helicobacter pylori induces caudal-type homeobox protein 2 and cyclooxygenase 2 expression by modulating microRNAs in esophageal epithelial cells. Cancer Sci 2018; 109(2):297–307.

79. Liu H, Li PW, Yang WQ, Mi H, Pan JL, Huang YC, *et al.* Identification of non-invasive biomarkers for chronic atrophic gastritis from serum exosomal microRNAs. BMC Cancer 2019; 19(1).

80. Wijnhoven BPL, Hussey DJ, Watson DJ, Tsykin A, Smith CM and Michael MZ. MicroRNA profiling of Barrett's oesophagus and oesophageal adenocarcinoma. Br J Surg 2010; 97(6):853–861.

81. Fassan M, Croce CM and Rugge M. miRNAs in precancerous lesions of the gastrointestinal tract. World J Gastroenterol 2011; 17(48):5231–5239.

82. Fassan M, Volinia S, Palatini J, Pizzi M, Baffa R, De Bernard M, *et al.* MicroRNA expression profiling in human Barrett's carcinogenesis. Int J cancer 2011; 129(7):1661–1670.

83. Bibi F, Naseer MI, Alvi SA, Yasir M, Jiman-Fatani AA, Sawan A, *et al.* microRNA analysis of gastric cancer patients from Saudi Arabian population. BMC Genomics 2016; 17(Suppl 9).

84. Yin C, Zheng X, Xiang H, Li H, Gao M, Meng X and Yang K. Differential expression profile analysis of cisplatin-regulated miRNAs in a human gastric cancer cell line. Mol Med Rep 2019; 20(2):1966–1976.

85. Wang D, Luo L and Guo J. miR-129-1-3p inhibits cell migration by targeting BDKRB2 in gastric cancer. Med Oncol 2014; 31(8).

86. Yu X, Luo L, Wu Y, Yu X, Liu Y, Yu X, *et al.* Gastric juice miR-129 as a potential biomarker for screening gastric cancer. Med Oncol 2013; 30(1).

87. Treece AL, Duncan DL, Tang W, Elmore S, Morgan DR, Dominguez RL, *et al.* Gastric adenocarcinoma microRNA profiles in fixed tissue and in plasma reveal cancer-associated and Epstein-Barr virus-related expression patterns. Lab Invest 2016; 96(6):661–671.

88. Sun M, Liu XH, Li JH, Yang JS, Zhang EB, Yin DD, *et al.* MiR-196a is upregulated in gastric cancer and promotes cell proliferation by downregulating p27(kip1). Mol Cancer Ther 2012; 11(4):842–852.

89. Ishiguro T, Ishiguro H, Kuwabara Y, Kimura M, Mitui A, Mori Y, *et al.* microRNA expression profile in undifferentiated gastric cancer. Int J Oncol 1992; 34(2):537–542.

90. Ueda T, Volinia S, Okumura H, Shimizu M, Taccioli C, Rossi S, *et al.* Relation between microRNA expression and progression and prognosis of gastric cancer: a microRNA expression analysis. Lancet Oncol 2010; 11(2):136–146.

91. Minna E, Romeo P, Dugo M, De Cecco L, Todoerti K, Pilotti S, *et al.* miR-451a is underexpressed and targets AKT/mTOR pathway in papillary thyroid carcinoma. Oncotarget 2016; 7(11):12731–47.

92. Sun H and Jiang P. MicroRNA-451a acts as tumor suppressor in cutaneous basal cell carcinoma. Mol Genet Genomic Med 2018; 6(6):1001–1009.

93. Yamada Y, Arai T, Sugawara S, Okato A, Kato M, Kojima S, *et al.* Impact of novel oncogenic pathways regulated by antitumor *miR-451a* in renal cell carcinoma. Cancer Sci 2018; 109(4):1239–1253.

94. El Bali L, Diman A, Bernard A, Roosens NHC and Dekeersmaecker SCJ. Comparative study of seven commercial kits for human DNA extraction from urine samples suitable for DNA biomarker-based public health studies. J Biomol Tech 2014; 25(4):96–110.

95. Chiarella P, Carbonari D, Capone P, Cavallo D, Iavicoli S, Mansi A, *et al.* Susceptibility biomarker detection in urine exfoliate DNA. Biomark Med 2017; 11(11):957–966.

96. El Messaoudi S, Rolet F, Mouliere F and Thierry AR. Circulating cell free DNA: Preanalytical considerations. Clin Chim Acta 2013; 424:222–230.

97. Raymond CK, Hernandez J, Karr R, Hill K and Li M. Collection of cell-free DNA for genomic analysis of solid tumors in a clinical laboratory setting. PLoS One 2017; 12(4):e0176241.

98. Sherwood JL, Corcoran C, Brown H, Sharpe AD, Musilova M and Kohlmann A. Optimised Pre-Analytical Methods Improve KRAS Mutation Detection in Circulating Tumour DNA (ctDNA) from Patients with Non-Small Cell Lung Cancer (NSCLC). PLoS One 2016; 11(2):e0150197.

99. Pérez-Barrios C, Nieto-Alcolado I, Torrente M, Jiménez-Sánchez C, Calvo V, Gutierrez-Sanz L, *et al.* Comparison of methods for circulating cell-free DNA isolation using blood from cancer patients: impact on biomarker testing. Transl lung cancer Res 2016; 5(6):665–672.

100. Medina Diaz I, Nocon A, Mehnert DH, Fredebohm J, Diehl F and Holtrup F. Performance of Streck cfDNA Blood Collection Tubes for Liquid Biopsy Testing. PLoS One 2016; 11(11):e0166354.

101. Kiyosawa N, Watanabe K, Toyama K and Ishizuka H. Circulating miRNA Signature as a Potential Biomarker for the Prediction of Analgesic Efficacy of Hydromorphone. Int J Mol Sci 2019, Vol 20, Page 1665 2019; 20(7):1665.

102. Lampignano R, Neumann MHD, Weber S, Kloten V, Herdean A, Voss T, *et al.* Multicenter evaluation of circulating cell-free DNA extraction and downstream analyses for the development of standardized (Pre)analytical work flows. Clin Chem 2020; 66(1):149–160.

103. Sai S, Ichikawa D, Tomita H, Ikoma D, Tani N, Ikoma H, *et al.* Quantification of plasma cell-free DNA in patients with gastric cancer. Anticancer Res 2007; 27(4 C):2747–2751.

104. Park JL, Kim HJ, Choi BY, Lee HC, Jang HR, Song KS, *et al.* Quantitative analysis of cell-free DNA in the plasma of gastric cancer patients. Oncol Lett 2012; 3(4):921–926.

105. Hsieh C-C, Hsu H-S, Chang S-C and Chen Y-J. Circulating Cell-Free DNA Levels Could Predict Oncological Outcomes of Patients Undergoing Esophagectomy for Esophageal Squamous Cell Carcinoma. Int J Mol Sci 2016; 17(12):2131.

106. Mouliere F, El Messaoudi S, Pang D, Dritschilo A and Thierry AR. Multi-marker analysis of circulating cell-free DNA toward personalized medicine for colorectal cancer. Mol Oncol 2014; 8(5):927–941.

107. Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, *et al.* COSMIC: The Catalogue Of Somatic Mutations In Cancer. Nucleic Acids Res 2019; 47(D1):D941–D947.

108. Kim YW, Kim YH, Song Y, Kim HS, Sim HW, Poojan S, *et al.* Monitoring circulating tumor DNA by analyzing personalized cancer-specific rearrangements to detect recurrence in gastric cancer. Exp Mol Med 2019; 51(8):1–10.

109. Lan J, Lu Y, Guan Y, Chang L, Yu Z and Qian H. Identification of circulating tumor DNA using a targeted 545–gene next generation sequencing panel in patients with gastric cancer. Oncol Lett 2020; 19(3):2251–2257.

110. Bu J, Hee Lee T, Jeong WJ, Poellmann MJ, Mudd K, Eun HS, *et al.* Enhanced detection of cell-free DNA (cfDNA) enables its use as a reliable biomarker for diagnosis and prognosis of gastric cancer. PLoS One 2020; 15(12):e0242145.

111. Ko K, Kananazawa Y, Yamada T, Kakinuma D, Matsuno K, Ando F, *et al.* Methylation status and long-fragment cell-free DNA are prognostic biomarkers for gastric cancer. Cancer Med 2021; 10(6):2003–2012.

112. Yang J, Gong Y, Lam VK, Shi Y, Guan Y, Zhang Y, *et al.* Deep sequencing of circulating tumor DNA detects molecular residual disease and predicts recurrence in gastric cancer. Cell Death Dis 2020; 11(5).

113. Yuan HL, Wang T and Zhang KH. MicroRNAs as potential biomarkers for diagnosis, therapy and prognosis of gastric cancer. Onco Targets Ther 2018; 11:3891.

SUPPLEMENTS

Supplement 1

KAUNO REGIONINIS BIOMEDICININIŲ TYRIMŲ ETIKOS KOMITETAS

KMUK Eivenių 2, Centrinis korpusas 71 kab., 50009 Kaunas, tel. +370 37 326168; faks. +370 37 326901, e-mail: cmeinfo@kmu.lt

LEIDIMAS ATLIKTI BIOMEDICININĮ TYRIMĄ

2011- 03 -08   Nr. BE-2- 10

Biomedicininio tyrimo pavadinimas:	**„Virškinimo sistemos ligų tiriamosios medžiagos biobankas".**
Protokolo Nr.:	**1**
Data:	**2010-12-27**
Versija:	**1**
Pagrindinis tyrėjas:	**Prof. habil. dr. Limas Kupčinskas** **Prof. habil. dr. Juozas Pundzius**
Biomedicininio tyrimo vieta:	**LSMU MA Gastroenterologijos klinika**
Įstaigos pavadinimas:	**LSMU MA Chirurgijos klinika**
Adresas:	**Eivenių g. 2, LT-50009 Kaunas**

Išvada:

Kauno regioninio biomedicininių tyrimų etikos komiteto posėdžio, įvykusio **2011 m. sausio 4 d.** (protokolo Nr. 8/2011) sprendimu pritarta biomedicininio tyrimo vykdymui.

Mokslinio eksperimento vykdytojai įsipareigoja: (1) nedelsiant informuoti Kauno Regioninį biomedicininių Tyrimų Etikos komitetą apie visus nenumatytus atvejus, susijusius su studijos vykdymu, (2) iki sausio 15 dienos – pateikti metinį studijos vykdymo apibendrinimą bei, (3) per mėnesį po studijos užbaigimo, pateikti galutinį pranešimą apie eksperimentą.

Kauno regioninio biomedicininių tyrimų etikos komiteto nariai			
Nr.	Vardas, Pavardė	Veiklos sritis	Dalyvavo posėdyje
1.	Doc. Irena Marchertienė	anesteziologija	taip
2.	Doc. Romaldas Mačiulaitis	klinikinė farmakologija	taip
3.	Prof. Nijolė Dalia Bakšienė	pediatrija	taip
4.	Prof. Irayda Jakušovaitė	filosofija	ne
5.	Dr.Eimantas Peičius	filosofija	taip
6.	Laima Vasiliauskaitė	psichoterapija	taip
7.	Gintaras Česnauskas	chirurgija	ne
8.	Zelmanas Šapiro	terapija	ne
9.	Jurgita Laurinaitytė	bioteisė	ne

Kauno regioninis biomedicininių tyrimų etikos komitetas dirba vadovaudamasis etikos principais nustatytais biomedicininių tyrimų Etikos įstatyme, Helsinkio deklaracijoje, vaistų tyrinėjimo Geros klinikinės praktikos taisyklėmis.

Pirmininkė   Irena Marchertienė

KAUNO REGIONINIS BIOMEDICININIŲ TYRIMŲ ETIKOS KOMITETAS

Lietuvos sveikatos mokslų universitetas, A. Mickevičiaus g. 9, LT 44307 Kaunas., tel. +370 37 32 68 89; e-mail: kaunorbtek@lsmuni.lt

PRITARIMAS
BIOMEDICININIO TYRIMO PAPILDYMUI/PRATĘSIMUI

2018-06-05 Nr. P1-BE-2-31/2018

Biomedicininio tyrimo pavadinimas: „**Daugiapakopis molekulinis virškinimo sistemos ligų tyrimas: genetinių, proteominių, epigenetinių ir mikrobiomo žymenų nustatymas**" (leidimo atlikti biomedicininį tyrimą leidimas 2018.03.22 Nr. BE-2-31).	
Pagrindinis tyrėjas:	**Prof.dr. Limas Kupčinskas**
Biomedicininio tyrimo vieta:	**Lietuvos sveikatos mokslų universiteto ligoninė Kauno klinikos** Eivenių g. 2, LT-50161, Kaunas.

Peržiūrėti šie [√] su minėtu tyrimu susiję dokumentai:

[√] Biomedicininio tyrimo: „Daugiapakopis molekulinis virškinimo sistemos ligų tyrimas: genetinių, proteominių, epigenetinių ir mikrobiomo žymenų nustatymas" pagrindinio tyrėjo prašymas dėl biomedicininio tyrimo papildymo/pakeitimo.

[√] Biomedicininio tyrimo: „Daugiapakopis molekulinis virškinimo sistemos ligų tyrimas: genetinių, proteominių, epigenetinių ir mikrobiomo žymenų nustatymas" paraiška.

[√] Biomedicininio tyrimo: „Daugiapakopis molekulinis virškinimo sistemos ligų tyrimas: genetinių, proteominių, epigenetinių ir mikrobiomo žymenų nustatymas" tyrėjo Dainiaus Jančiausko gyvenimo aprašymas.

[√] Nutarta:
[√] Pritarti biomedicininio tyrimo papildymui/pakeitimui.

	Kauno regioninio biomedicininių tyrimų etikos komiteto nariai	
Nr.	Vardas, Pavardė	Veiklos sritis
1.	Prof. Edgaras Stankevičius	Fiziologija, farmakologija
2.	Prof. Skaidrius Miliauskas	Chirurgija
3.	Prof. Kęstutis Petrikonis	Neurologija
4.	Doc. Gintautas Gumbrevičius	Klinikinė farmakologija
5.	Med. dr. Jonas Andriuškevičius	Chirurgija
6.	Dr. Ramunė Kasperavičienė	Kalbotyra
7.	Eglė Vaižgelienė	Visuomenės sveikata
8.	Žydrūnė Luneckaitė	Filosofija
9.	Jurgita Laurinaitytė	Teisė

Kauno regioninis biomedicininių tyrimų etikos komitetas dirba vadovaudamasis etikos principais nustatytais biomedicininių tyrimų Etikos įstatyme, Helsinkio deklaracijoje, vaistų tyrinėjimo Geros klinikinės praktikos taisyklėmis.

Pirmininkas   Prof. Edgaras Stankevičius

MEDIZINISCHE FAKULTÄT
DER CHRISTIAN-ALBRECHTS-UNIVERSITÄT ZU KIEL

ETHIK-KOMMISSION

A.D. 1665

Universitäts-Kinderklinik · Schwanenweg 20 · 24105 Kiel

Prof. Dr. rer. nat. Norbert Arnold
Onkologisches Labor
UKSH, Campus Kiel
Arnold-Heller-Str. 3, Haus 24
24105 Kiel

Schwanenweg 20
D-24105 Kiel

Telefon 04 31 / 597-18 09
Telefax 04 31 / 597-18 31

Datum: 11.01.2011

AZ - Kiel:	**B 327/10** (bitte stets angeben)
Studienplan:	**Consortium Mammakarzinom Tumorbank**
	Studienplan, Patienteninformation und Einwilligungserklärung
	Voten der Ethik-Kommission der Albert-Ludwigs-Universität Freiburg vom 27.11.2007 und der Ethik-Kommission der Medizinischen Fakultät der E-berhard-Karls-Universität Tübingen vom 03.03.1999 und 06.10.2005, Bewilligungsbescheid der Deutschen Krebshilfe vom 27.11.2009
Projektleiter:	**Prof. Dr. Axel zur Hausen, Institut für Pathologie, Freiburg**
Datum des Antrages:	**15.12.2010**

Sehr geehrter Herr Arnold,
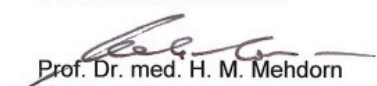
vielen Dank für Ihren obengenannten Antrag zur Beratung gemäß § 15 Berufsordnung (BO) der Ärztekammer Schleswig-Holstein.

Nach Durchsicht der Unterlagen durch die Geschäftsstelle und durch mich als Vorsitzenden der Ethik-Kommission bestehen gegen das Projekt keine berufsethischen und berufsrechtlichen Bedenken.

Die im Folgenden aufgeführten Hinweise sollten jedoch beachtet werden:

1. Die Ethik-Kommission empfiehlt eine Änderung der Überschrift der Patienten-information und Einverständniserklärung: es handelt sich nicht um eine „Zustimmung zur Gewebeentnahme", sondern um eine Zustimmung zur Aufbewahrung und wissenschaftlichen Verwendung von Rest-Gewebeproben und Körperflüssigkeiten.

2. In der Patienteninformation und Einverständniserklärung sollte der Name und die Tel.-Nr. des verantwortlichen Arztes für eventuelle Rückfragen aufgeführt werden.

Mit freundlichen Grüßen

Prof. Dr. med. H. M. Mehdorn
Vorsitzender der Ethik-Kommission

Dr. med. Christine Glinicke
Geschäftsführung der Ethik-Kommission

**MEDIZINISCHE FAKULTÄT
DER CHRISTIAN-ALBRECHTS-UNIVERSITÄT ZU KIEL**

ETHIK-KOMMISSION

A.D. 1665

Universitäts-Kinderklinik · Schwanenweg 20 · 24105 Kiel

Prof. Dr. rer. nat. Burkhard Brandt
Institut für Klinische Chemie
Arnold-Heller-Straße 3, Haus 17
24105 Kiel

Postadresse:
Arnold-Heller-Straße 3 / Haus 9
D-24105 Kiel

Telefon   04 31 / 597-18 09
Telefax   04 31 / 597-53 33
ethikkomm@email.uni-kiel.de

Datum: 17. Februar 2015

AZ.:	**D 470/14** (bitte stets angeben)
	In Bezug auf AZ.:B 327/10
Studienplan:	**Mammakarzinom - Zusätzliche prognostische und prädiktive diagnostische Kriterien werden durch einen Nachweis von Tumorzellen im Blut (TIB) gewonnen**
	Anschreiben, Projektbeschreibung, Patienteninformation, Version 1.0 vom 19.1.2011 Gewebebank der Klinik für Gynäkologie und Geburtshilfe in Zusammenarbeit mit dem Institut für Pathologie des UKSH, Campus Kiel
	Einverständniserklärung Version 1.0 vom 19.1.2011 zur Aufbewahrung und Weiterverarbeitung von Gewebeproben und Körperflüssigkeiten in der Gewebebank
Kooperation mit:	**PD Dr. med. Schem, Klinik für Gynäkologie und Geburtshilfe, UKSH Campus Kiel**
Ihre Email vom: Nachreichung /	**23. Juli 2014 (Eingang: 25. Juli 2014)**
Überarbeitung vom:	**4. Februar 2015 (Eingang: 6. Februar 2015)**
	Punkt-für-Punkt-Antwort auf Schreiben der Ethik-Kommission vom 31. Juli 2014 von Prof. N. Arnold (Klinik für Gynäkologie und Geburtshilfe); Einverständniserklärung von Prof. Jonat vom 4. Februar 2015; Patienteninformation Gewebebank vom 4. Februar 2015; Einverständniserklärung Gewebebank vom 4. Februar 2015

Sehr geehrter Herr Professor Brandt,

wir bestätigen den Eingang des obengenannten Antrages zur Beratung gemäß § 15 Berufsordnung (BO) der Ärztekammer Schleswig-Holstein. Nach Durchsicht der Unterlagen durch die Geschäftsstelle und durch mich als stellvertretenden Vorsitzenden der Ethik-Kommission bestehen gegen die Durchführung der Studie nunmehr keine berufsethischen und berufsrechtlichen Bedenken.
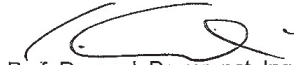
Die im Folgenden aufgeführten Hinweise müssen beachtet werden:
1. Es wird darauf hingewiesen, dass künftige Änderungen der Studie der Ethik-Kommission anzuzeigen sind und gegebenenfalls eine erneute Beratung erforderlich machen.

2. Die ethische und rechtliche Verantwortung für die Durchführung dieser Studie verbleibt beim Projektleiter und den an der Studie teilnehmenden Ärzten.

3. Die Ethik-Kommission weist darauf hin, dass für eventuell in Zukunft weitere teilnehmende Zentren eine berufsrechtliche Beratung bei der jeweils für sie zuständigen Ethik-Kommission erforderlich ist.

4. Gemäß Deklaration von Helsinki **muss** der Ethik-Kommission nach Studienende ein

Abschlussbericht vorgelegt werden, der eine Zusammenfassung der Ergebnisse und Schlussfolgerungen der Studie enthält.

Mit freundlichen kollegialen Grüßen

Prof. Dr. med. Dr. rer. nat. Ingolf Cascorbi
Stv. Vorsitzender der Ethik-Kommission

Dr. med. Christine Glinicke
Geschäftsführung der Ethik-Kommission

CCC | Marketplace™

Order Date	21-Feb-2022	Type of Use	Republish in a thesis/dissertation
Order License ID	1191309-1	Publisher	Mary Ann Liebert, Inc.
ISSN	1947-5543	Portion	Chapter/article

LICENSED CONTENT

Publication Title	Biopreservation and biobanking	Publication Type	e-Journal
Article Title	Effects of Quantification Methods, Isolation Kits, Plasma Biobanking, and Hemolysis on Cell-Free DNA Analysis in Plasma.	Start Page	553
		End Page	561
Author/Editor	International Society for Biological and Environmental Repositories.	Issue	6
		Volume	17
Date	01/01/2009	URL	http://www.liebertpub.com/cpt
Language	English		
Country	United States of America		
Rightsholder	Mary Ann Liebert Inc.		

REQUEST DETAILS

Portion Type	Chapter/article	Rights Requested	Main product
Page range(s)	553-561	Distribution	Worldwide
Total number of pages	11	Translation	Original language of publication
Format (select all that apply)	Print, Electronic	Copies for the disabled?	No
Who will republish the content?	Academic institution	Minor editing privileges?	No
Duration of Use	Life of current and all future editions	Incidental promotional use?	No
Lifetime Unit Quantity	Up to 499	Currency	EUR

NEW WORK DETAILS

Title	Comprehensive analysis of circulating nucleic acids in gastric cancer patients' plasma	Institution name	Lithuanian University of Health Sciences
		Expected presentation date	2022-07-04
Instructor name	Dissertation supervisor dr. Jurgita Skieceviciene		

ADDITIONAL DETAILS

Order reference number	N/A	The requesting person / organization to appear on the license	Greta Varkalaite

REUSE CONTENT DETAILS

Title, description or numeric reference of the portion(s)	Whole publication	Title of the article/chapter the portion is from	Effects of Quantification Methods, Isolation Kits, Plasma Biobanking, and Hemolysis on Cell-Free DNA Analysis in Plasma.
Editor of portion(s)	Streleckiene, Greta; Forster, Michael; Inciuraite, Ruta; Lukosevicius, Rokas; Skieceviciene, Jurgita	Author of portion(s)	Streleckiene, Greta; Forster, Michael; Inciuraite, Ruta; Lukosevicius, Rokas; Skieceviciene, Jurgita
Volume of serial or monograph	17		
Page or page range of portion	553-561	Issue, if republishing an article from a serial	6
		Publication date of portion	2019-12-11

RIGHTSHOLDER TERMS AND CONDITIONS

CCC Terms and Conditions

3.2. General Payment Terms: You may pay by credit card or through an account with us payable at the end of the month. If you and we agree that you may establish a standing account with CCC, then the following terms apply: Remit Payment to: Copyright Clearance Center, 29118 Network Place, Chicago, IL 60673-1291. Payments Due: Invoices are payable upon their delivery to you (or upon our notice to you that they are available to you for downloading). After 30 days, outstanding amounts will be subject to a service charge of 1-1/2% per month or, if less, the maximum rate allowed by applicable law. Unless otherwise specifically set forth in the Order Confirmation or in a separate written agreement signed by CCC, invoices are due and payable on "net 30" terms. While User may exercise the rights licensed immediately upon issuance of the Order Confirmation, the license is automatically revoked and is null and void, as if it had never been issued, if complete payment for the license is not received on a timely basis either from User directly or through a payment agent, such as a credit card company.

3.3. Unless otherwise provided in the Order Confirmation, any grant of rights to User (i) is "one-time" (including the editions and product family specified in the license), (ii) is non-exclusive and non-transferable and (iii) is subject to any and all limitations and restrictions (such as, but not limited to, limitations on duration of use or circulation) included in the Order Confirmation or invoice and/or in these terms and conditions. Upon completion of the licensed use, User shall either secure a new permission for further use of the Work(s) or immediately cease any new use of the Work(s) and shall render inaccessible (such as by deleting or by removing or severing links or other locators) any further copies of the Work (except for copies printed on paper in accordance with this license and still in User's stock at the end of such period).

3.4. In the event that the material for which a republication license is sought includes third party materials (such as photographs, illustrations, graphs, inserts and similar materials) which are identified in such material as having been used by permission, User is responsible for identifying, and seeking separate licenses (under this Service or otherwise) for, any of such third party materials; without a separate license, such third party materials may not be used.

3.5. Use of proper copyright notice for a Work is required as a condition of any license granted under the Service. Unless otherwise provided in the Order Confirmation, a proper copyright notice will read substantially as follows: "Republished with permission of [Rightsholder's name], from [Work's title, author, volume, edition number and year of copyright]; permission conveyed through Copyright Clearance Center, Inc. " Such notice must be provided in a reasonably legible font size and must be placed either immediately adjacent to the Work as used (for example, as part of a by-line or footnote but not as a separate electronic link) or in the place where substantially all other credits or notices for the new work containing the republished Work are located. Failure to include the required notice results in loss to the Rightsholder and CCC, and the User shall be liable to pay liquidated damages for each such failure equal to twice the use fee specified in the Order Confirmation, in addition to the use fee itself and any other fees and charges specified.

3.6. User may only make alterations to the Work if and as expressly set forth in the Order Confirmation. No Work may be used in any way that is defamatory, violates the rights of third parties (including such third parties' rights of copyright, privacy, publicity, or other tangible or intangible property), or is otherwise illegal, sexually explicit or obscene. In addition, User may not conjoin a Work with any other material that may result in damage to the reputation of the Rightsholder. User agrees to inform CCC if it becomes aware of any infringement of any rights in a Work and to cooperate with any reasonable request of CCC or the Rightsholder in connection therewith.

4. Indemnity. User hereby indemnifies and agrees to defend the Rightsholder and CCC, and their respective employees and directors, against all claims, liability, damages, costs and expenses, including legal fees and expenses, arising out of any use of a Work beyond the scope of the rights granted herein, or any use of a Work which has been altered in any unauthorized way by User, including claims of defamation or infringement of rights of copyright, publicity, privacy or other tangible or intangible property.

5. Limitation of Liability. UNDER NO CIRCUMSTANCES WILL CCC OR THE RIGHTSHOLDER BE LIABLE FOR ANY DIRECT, INDIRECT, CONSEQUENTIAL OR INCIDENTAL DAMAGES (INCLUDING WITHOUT LIMITATION DAMAGES FOR LOSS OF BUSINESS PROFITS OR INFORMATION, OR FOR BUSINESS INTERRUPTION) ARISING OUT OF THE USE OR INABILITY TO USE A WORK, EVEN IF ONE OF THEM HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. In any event, the total liability of the Rightsholder and CCC (including their respective employees and directors) shall not exceed the total amount actually paid by User for this license. User assumes full liability for the actions and omissions of its principals, employees, agents, affiliates, successors and assigns.

6. Limited Warranties. THE WORK(S) AND RIGHT(S) ARE PROVIDED "AS IS". CCC HAS THE RIGHT TO GRANT TO USER THE RIGHTS GRANTED IN THE ORDER CONFIRMATION DOCUMENT. CCC AND THE RIGHTSHOLDER DISCLAIM ALL OTHER WARRANTIES RELATING TO THE WORK(S) AND RIGHT(S), EITHER EXPRESS OR IMPLIED, INCLUDING WITHOUT LIMITATION IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. ADDITIONAL RIGHTS MAY BE REQUIRED TO USE ILLUSTRATIONS, GRAPHS, PHOTOGRAPHS, ABSTRACTS, INSERTS OR OTHER PORTIONS OF THE WORK (AS OPPOSED TO THE ENTIRE WORK) IN A MANNER CONTEMPLATED BY USER; USER UNDERSTANDS AND AGREES THAT NEITHER CCC NOR THE RIGHTSHOLDER MAY HAVE SUCH ADDITIONAL RIGHTS TO GRANT.

7. Effect of Breach. Any failure by User to pay any amount when due, or any use by User of a Work beyond the scope of the license set forth in the Order Confirmation and/or these terms and conditions, shall be a material breach of the license created by the Order Confirmation and these terms and conditions. Any breach not cured within 30 days of written notice thereof shall result in immediate termination of such license without further notice. Any unauthorized (but licensable) use of a Work that is terminated immediately upon notice thereof may be liquidated by payment of the Rightsholder's ordinary license price therefor; any unauthorized (and unlicensable) use that is not terminated immediately for any reason (including, for example, because materials containing the Work cannot reasonably be recalled) will be subject to all remedies available at law or in equity, but in no event to a payment of less than three times the Rightsholder's ordinary license price for the most closely analogous licensable use plus Rightsholder's and/or CCC's costs and expenses incurred in collecting such payment.

8. Miscellaneous.

8.1. User acknowledges that CCC may, from time to time, make changes or additions to the Service or to these terms and conditions, and CCC reserves the right to send notice to the User by electronic mail or otherwise for the purposes of notifying User of such changes or additions; provided that any such changes or additions shall not apply to permissions already secured and paid for.

8.2. Use of User-related information collected through the Service is governed by CCC's privacy policy, available online here:https://marketplace.copyright.com/rs-ui-web/mp/privacy-policy

8.3. The licensing transaction described in the Order Confirmation is personal to User. Therefore, User may not assign or transfer to any other person (whether a natural person or an organization of any kind) the license created by the Order Confirmation and these terms and conditions or any rights granted hereunder; provided, however, that User may assign such license in its entirety on written notice to CCC in the event of a transfer of all or substantially all of User's rights in the new material which includes the Work(s) licensed under this Service.

8.4. No amendment or waiver of any terms is binding unless set forth in writing and signed by the parties. The Rightsholder and CCC hereby object to any terms contained in any writing prepared by the User or its principals, employees, agents or affiliates and purporting to govern or otherwise relate to the licensing transaction described in the Order Confirmation, which terms are in any way inconsistent with any terms set forth in the Order Confirmation and/or in these terms and conditions or CCC's standard operating procedures, whether such writing is prepared prior to, simultaneously with or subsequent to the Order Confirmation, and whether such writing appears on a copy of the Order Confirmation or in a separate instrument.

8.5. The licensing transaction described in the Order Confirmation document shall be governed by and construed under the law of the State of New York, USA, without regard to the principles thereof of conflicts of law. Any case, controversy, suit, action, or proceeding arising out of, in connection with, or related to such licensing transaction shall be brought, at CCC's sole discretion, in any federal or state court located in the County of New York, State of New York, USA, or in any federal or state court whose geographical jurisdiction covers the location of the Rightsholder set forth in the Order Confirmation. The parties expressly submit to the personal jurisdiction and venue of each such federal or state court.If you have any comments or questions about the Service or Copyright Clearance Center, please contact us at 978-750-8400 or send an e-mail to support@copyright.com.

v 1.1

COPIES OF PUBLICATIONS

A1

Title: Atrophic gastritis and gastric cancer tissue miRNome analysis reveal hsa-miR-129-1 and hsa-miR-196a as potential early diagnostic biomarkers

Authors: Varkalaitė, Greta; Vaitkevičiūtė, Evelina; Inčiūraitė, Rūta; Šaltenienė, Violeta; Juzėnas, Simonas; Petkevičius, Vytenis; Gudaitytė, Rita; Mickevičius, Antanas; Link, Alexander; Kupčinskas, Limas; Leja, Mārcis; Kupčinskas, Juozas; Skiecevičienė, Jurgita

World journal of gastroenterology 28, 6 (2022)

Reprinted under a Creative Commons Attribution-Non-Commercial (CC BY-NC 4.0) Open-Access licence

ORIGINAL ARTICLE

Case Control Study

Atrophic gastritis and gastric cancer tissue miRNome analysis reveals hsa-miR-129-1 and hsa-miR-196a as potential early diagnostic biomarkers

Greta Varkalaite, Evelina Vaitkeviciute, Ruta Inciuraite, Violeta Salteniene, Simonas Juzenas, Vytenis Petkevicius, Rita Gudaityte, Antanas Mickevicius, Alexander Link, Limas Kupcinskas, Marcis Leja, Juozas Kupcinskas, Jurgita Skieceviciene

ORCID number: Greta Varkalaite 0000-0003-3488-2171; Evelina Vaitkeviciute 0000-0003-2525-1597; Ruta Inciuraite 0000-0002-5806-6155; Violeta Salteniene 0000-0002-7881-945X; Simonas Juzenas 0000-0001-9293-0691; Vytenis Petkevicius 0000-0002-1964-5226; Rita Gudaityte 0000-0001-7489-8069; Antanas Mickevicius 0000-0002-6208-7039; Alexander Link 0000-0002-9514-4562; Limas Kupčinskas 0000-0002-8689-9023; Marcis Leja 0000-0002-0319-8855; Juozas Kupcinskas 0000-0002-8760-7416; Jurgita Skieceviciene 0000-0002-4893-6612.

Author contributions: Varkalaite G, Leja M, Kupcinskas L, Kupcinskas J and Skieceviciene J contributed to study conception and designed the research; Gudaityte R, Petkevicius V and Mickevicius A collected material and clinical data from the participants; Varkalaite G and Vaitkeviciute E performed all the investigations, analysis of data and drafted the manuscript; Juzenas S and Inciuraite R reviewed and edited the manuscript; Salteniene V revised the article for important intellectual content; Link A, Kupcinskas J and Skieceviciene J approved the final version of the manuscript.

Greta Varkalaite, Evelina Vaitkeviciute, Ruta Inciuraite, Violeta Salteniene, Simonas Juzenas, Limas Kupcinskas, Jurgita Skieceviciene, Institute for Digestive Research, Lithuanian University of Health Sciences, Kaunas 44307, Lithuania

Vytenis Petkevicius, Juozas Kupcinskas, Department of Gastroenterology, Lithuanian University of Health Sciences, Kaunas 44307, Lithuania

Rita Gudaityte, Antanas Mickevicius, Department of Surgery, Lithuanian University of Health Sciences, Kaunas 44307, Lithuania

Alexander Link, Department of Gastroenterology, Hepatology and Infectious Diseases, Otto-von-Guericke University Hospital, Magdeburg 39120, Germany

Marcis Leja, Institute of Clinical and Preventive Medicine & Faculty of Medicine, University of Latvia, Riga 1586, Latvia

Corresponding author: Jurgita Skieceviciene, PhD, Senior Researcher, Institute for Digestive Research, Lithuanian University of Health Sciences, A. Mickeviciaus street 9, Kaunas 44307, Lithuania. jurgita.skieceviciene@lsmuni.lt

Abstract

BACKGROUND

Gastric cancer (GC) is one of the most frequently diagnosed tumor globally. In most cases, GC develops in a stepwise manner from chronic gastritis or atrophic gastritis (AG) to cancer. One of the major issues in clinical settings of GC is diagnosis at advanced disease stages resulting in poor prognosis. MicroRNAs (miRNAs) are small noncoding molecules that play an essential role in a variety of fundamental biological processes. However, clinical potential of miRNA profiling in the gastric cancerogenesis, especially in premalignant GC cases, remains unclear.

AIM

To evaluate the AG and GC tissue miRNomes and identify specific miRNAs' potential for clinical applications (*e.g.*, non-invasive diagnostics).

77

METHODS

Study included a total of 125 subjects: Controls (CON), AG, and GC patients. All study subjects were recruited at the Departments of Surgery or Gastroenterology, Hospital of Lithuanian University of Health Sciences and divided into the profiling (n = 60) and validation (n = 65) cohorts. Total RNA isolated from tissue samples was used for preparation of small RNA sequencing libraries and profiled using next-generation sequencing (NGS). Based on NGS data, deregulated miRNAs hsa-miR-129-1-3p and hsa-miR-196a-5p were analyzed in plasma samples of independent cohort consisting of CON, AG, and GC patients. Expression level of hsa-miR-129-1-3p and hsa-miR-196a-5p was determined using the quantitative real-time polymerase chain reaction and $2^{-\Delta\Delta Ct}$ method.

RESULTS

Results of tissue analysis revealed 20 differentially expressed miRNAs in AG group compared to CON group, 129 deregulated miRNAs in GC compared to CON, and 99 altered miRNAs comparing GC and AG groups. Only 2 miRNAs (hsa-miR-129-1-3p and hsa-miR-196a-5p) were identified to be step-wise deregulated in healthy-premalignant-malignant sequence. Area under the curve (AUC)-receiver operating characteristic analysis revealed that expression level of hsa-miR-196a-5p is significant for discrimination of CON *vs* AG, CON *vs* GC and AG *vs* GC and resulted in AUCs: 88.0%, 93.1% and 66.3%, respectively. Comparing results in tissue and plasma samples, hsa-miR-129-1-3p was significantly down-regulated in GC compared to AG (P = 0.0021 and P = 0.024, tissue and plasma, respectively). Moreover, analysis revealed that hsa-miR-215-3p/5p and hsa-miR-934 were significantly deregulated in GC based on *Helicobacter pylori* (*H. pylori*) infection status [log2 fold change (FC) = -4.52, P-adjusted = 0.02; log2FC = -4.00, P-adjusted = 0.02; log2FC = 6.09, P-adjusted = 0.02, respectively].

CONCLUSION

Comprehensive miRNome study provides evidence for gradual deregulation of hsa-miR-196a-5p and hsa-miR-129-1-3p in gastric carcinogenesis and found hsa-miR-215-3p/5p and hsa-miR-934 to be significantly deregulated in *H. pylori* carrying GC patients.

Key Words: Gastric cancer; Atrophic gastritis; Tumorigenesis; *Helicobacter pylori*; MicroRNAs; Biomarkers

Core Tip: In this research we aimed to evaluate microRNAs profiles of premalignant and malignant stages of gastric cancer (GC). To date this is the first study analyzing atrophic gastritis (AG) and GC tissue miRNomes in the subjects of European origin using next-generation sequencing approach. We showed that hsa-miR-196a-5p expression in tissue is significant for discrimination between controls and AG or GC, while hsa-miR-129-1-3p is potential candidate for non-invasive GC diagnostic. This study provides novel insights into complex GC pathogenesis cascade and might be highly significant for future studies of new AG or GC associated epigenetic markers or even diagnostic targets.

INTRODUCTION

Gastric cancer (GC) is the one of the most common malignancy and the fourth leading cause of cancer-related death worldwide[1]. Studies show, that in most cases GC development is a stepwise process: Chronic gastric mucosa inflammation progresses to atrophic gastritis (AG) or intestinal metaplasia (IM), which eventually may become predisposition to GC. This complex cascade involves many factors: *Helicobacter pylori* (*H. pylori*) infection, lifestyle, dietary habits, and genetic or epigenetic alterations, including miRNA expression changes[2,3]. One of the major concerns in diagnostics of GC is poor survival rate and prognosis, while this tumor is usually diagnosed at late stages. Therefore, investigation of the molecular mechanisms that are critical in the complex GC pathological cascade may help to identify novel therapeutic targets and consequently improve the disease prognosis. MicroRNAs (miRNAs) are small (approx 22 nt) non-coding RNA molecules that regulate gene expression by binding to the specific sites within 3' untranslated regions of target mRNAs[4,5]. MiRNAs play very important role in many physiological and pathological processes as well as tumori-genesis and may function as either tumor-suppressors or as oncogenic miRNAs[6-8]. Studies have reported numerous differentially expressed miRNAs in malignant gastric tissues including members of miR-20, miR-451, miR-148, miR-223 families[9-11]. Despite the previous efforts and conducted miRNA studies in GC, the miRNome characterization of premalignant gastric condition - AG - remains largely unknown.

In this study, we aimed to investigate miRNome profile through GC tumorigenesis cascade including precancerous lesions, such as AG. Also, expression of two miRNAs (hsa-miR-129-1 and hsa-miR-196a) was analyzed in plasma samples of the independent cohort of AG and GC patients. Tissue miRNome analysis results revealed distinct miRNA profiles comparing controls (CON), AG, and GC groups. Also, our study findings show that two miRNAs: Hsa-miR-129-1 and hsa-miR-196a may be a relevant biomarker for GC diagnostics.

MATERIALS AND METHODS

Study population

The study included a total of 125 CON and patients diagnosed with AG and GC, who were divided into the profiling cohort of 60 subjects and validation cohort of 65 subjects. Tissue samples of profiling cohort were collected during the years 2007-2015, while plasma of participants in validation cohort was collected from years 2011-2019 at the Departments of Surgery and Gastroenterology, Hospital of Lithuanian University of Health Sciences (Kaunas, Lithuania). Clinical and phenotypic characteristics of subjects investigated in profiling and validation cohorts are presented in Table 1. *H. pylori* status was assessed using indirect ELISA to detect serum-specific IgG antigen (Virion/Serion GmbH, Germany). Control group consisted of subjects, who had no signs of atrophy or IM according to Operative Link on Gastritis Assessment (OLGA) staging system (stage 0)[12]. AG group consisted of individuals that had stage I-IV atrophy score in gastric mucosa by OLGA classification. Gastric adenocarcinoma in GC patients was verified by histology and classified according to the American Joint Committee on Cancer TNM Staging Classification and Lauren Classification[13,14]. Adjacent GC (GCaj) samples were biopsy samples obtained from endoscopically healthy appearing gastric mucosa at least 2 cm away from the primary tumor.

The study was approved by the Kaunas Regional Biomedical Research Ethics Committee (approval No BE-2-10 and BE-2-31) and performed in accordance with the Declaration of Helsinki. All study participants provided written informed consent before enrollment.

Total RNA extraction

Total RNA, including small RNA fraction, was isolated from CON, AG and GC tissues using miRNeasy Mini Kit (Qiagen, Germany) according to the manufacturer's instructions. Quantification of RNA was performed using Nanodrop2000 spectropho-tometer (Thermo Fisher Scientific, United States) and quality of RNA samples was evaluated by Agilent 2100 Bioanalyzer (Agilent Technologies, United States). Circulating nucleic acids, including circulating miRNA fraction, was isolated using QIAamp Circulating Nucleic Acid Kit (Qiagen, Germany) according to manufacturer's instructions. All isolated samples were stored at -80 °C prior to further analysis.

79

Table 1 Demographic characteristics of profiling and validation cohorts						
		Profiling cohort (*n* = 60)			Validation cohort (*n* = 65)		
		CON (*n* = 21)	AG (*n* = 19)	GC (*n* = 20)	CON (*n* = 11)	AG (*n* = 30)	GC (*n* = 24)
Age	Mean ± SD	58.29 ± 15.52	69.21 ± 8.78	64.95 ± 10.89	42.27 ± 12.89	68.01 ± 11.81	68.33 ± 11.27
Gender (*n*)	Male	5	3	15	5	9	18
	Female	16	16	5	6	21	6
Helicobacter pylori infection (*n*)	Negative	12	10	8	-	17	9
	Positive	9	9	9	-	10	4
	Unknown	-	-	3	11	3	11
Differentiation grade (*n*)	G1	-	-	4	-	-	-
	G2	-	-	4	-	-	12
	G3	-	-	12	-	-	12
Lauren classification (*n*)	Diffuse	-	-	10	-	-	8
	Intestinal	-	-	10	-	-	13
	Mixed	-	-	-	-	-	2
	Unknown	-	-	-	-	-	1
T (*n*)	T1	-	-	6	-	-	3
	T2	-	-	2	-	-	5
	T3	-	-	8	-	-	9
	T4	-	-	4	-	-	6
	Unknown	-	-	-	-	-	1
N (*n*)	N0	-	-	10	-	-	6
	N1	-	-	2	-	-	5
	N2	-	-	3	-	-	4
	N3	-	-	5	-	-	8
	Unknown	-	-	-	-	-	1
M (*n*)	M0	-	-	7	-	-	14
	M1	-	-	2	-	-	9
	Unknown	-	-	11	-	-	1

SD: Standard deviation; CON: Control; AG: Atrophic gastritis; GC: Gastric cancer.

Small RNA-seq library preparation and next-generation sequencing

Small RNA libraries were prepared using Illumina TruSeq Small RNA Sample Preparation Kit (Illumina, United States) according to the manufacturer's protocol with 1 μg RNA input per sample followed by RNA 3′ adapter ligation, RNA 5′ adapter ligation, cDNA synthesis, polymerase chain reaction (PCR) amplification using unique barcode sequences for each sample and gel size-selection of small RNA library. The yield and quality of sequencing libraries were assessed using the Agilent 2100 Bioanalyzer (Agilent Technologies, United States). The small RNA libraries were randomized, pooled 24 samples per lane and sequenced using Illumina HiSeq 2500 (1 × 50 bp single-end reads).

Bioinformatics analysis of small RNA-seq data

Analysis of raw small RNA-seq data was performed by nf-core/smrnaseq pipeline v.1.0.0 including Nextflow v.20.07.1[15], Java v.11.0.7, and Docker v.19.03.12. In brief, all steps consisted of read quality control using FastQC v.0.11.9, removing 3′ adapter sequences with TrimGalore! v.0.6.5, mapping to mature and hairpin miRNAs (miRBase v.22.1[16]), and GRCh37 human reference genome with Bowtie v.1.3.0[17].

80

After alignment and trimming sorted BAM files were used for further analysis with edgeR v.3.32.1[18] and mirtop v.0.4.23. MiRNA quality was assessed and summarized using MultiQC v.1.9[19]. Normalized counts were generated using isomiRs package and differential expression analysis was carried out using the DESeq2 Bioconductor package v.1.26.0[20]. The threshold for significant differential expression was Bonferroni[21] adjusted *P*-value < 0.05 and absolute value of log2 fold change (FC) |log2FC| > 1.

Validation of miRNA expression in plasma by reverse transcription quantitative real-time PCR

To validate differentially expressed miRNAs in plasma samples, isolated plasma circulating microRNA was reverse transcribed to cDNA using the TaqMan™ MicroRNA Reverse Transcription Kit (Thermo Fisher Scientific, United States). The material was preamplified using the TaqMan PreAmp Master Mix (Applied Biosystems, United States) according to the manufacturer's protocol. Quantitative real-time PCR (RT-PCR) was performed using the TaqMan MicroRNA Assays: Hsa-miR-129* (Assay ID: 002298), hsa-miR-196a (Assay ID: 241070_mat) on 7500 Fast Real-Time PCR System (Applied Biosystems, United States). All RT-qPCR reactions were run in duplicate in a 20 µL reaction and the relative fold change in miRNA expression was estimated using the $2^{-\Delta\Delta Ct}$ method[22]. Ct values were normalized to the RNU6B (Assay ID: 001093, Thermo Fisher Scientific, United States) endogenous control.

Statistical analysis

Statistical analysis was performed using RStudio software (R v.3.6.3). Shapiro-Wilk normality test was used to test the normal distribution of data. For normally distributed data, statistical significance was assessed by Student's *t*-test. If the data did not pass normality tests was performed non-parametric Wilcoxon rank-sum test. A *P* < 0.05 was considered statistically significant. Area under the receiver operating characteristic curve (AUC-ROC) analysis was performed using pROC R package.

RESULTS

Small RNA sequencing reveals distinct miRNomes of healthy, premalignant, and malignant stages of GC

Small RNA sequencing of CON, AG, and paired GC (cancerous and adjacent) tissues in total identified 1037 miRNAs annotated in the miRBase v22.1. Sequencing yielded approx 250 M raw sequencing reads (from 359 K to 16 M reads per sample). After quality control steps 396 low-abundant and non-variable miRNAs and 5 outlying samples were removed resulting in 641 miRNAs and 75 samples which were used for further analysis (Supplementary Figures 1 and 2). The number of deregulated miRNAs corresponded to pathological cascade of GC development. The highest number of deregulated miRNAs were determined when comparing GC and CON groups (129 differentially expressed miRNAs, 82 up-regulated and 47 down-regulated; Supplementary Table 1). Next, 99 differentially expressed miRNAs were identified analyzing GC compared to AG (67 up-regulated and 32 down-regulated; Supplementary Table 2). The lowest number, 20 miRNAs, were found to be deregulated comparing AG and CON (6 up-regulated and 14 down-regulated; Supplementary Table 3). Differential expression results comparing GC *vs* GCaj, AG *vs* GCaj, and CON *vs* GCaj are presented in Supplementary Tables 4, 5 and 6 respectively.

Differential expression results and top five deregulated miRNAs in each case are represented in Figure 1A. Multidimensional scaling analysis of normalized expression values, assessing the similarity structure of miRNomes (Spearman's correlation distance), revealed 4 clusters, corresponding to the CON, AG, GC cancerous and adjacent tissues (Figure 1B). The AG cluster was intermediate between GC and CON, whereas GCaj was overlapping with AG and CON groups.

Hsa-miR-129-1-3p and hsa-miR-196a-5p may be employed for discrimination of healthy, premalignant, and malignant GC cases

To further study miRNome profiles, altered expression of miRNAs was analyzed in three main comparison groups: AG *vs* CON, GC *vs* CON and AG *vs* GC according to clinical significance. Analyzing uniquely deregulated miRNAs, 40 differentially expressed miRNAs were found when compared GC to CON (25.8% of all deregulated
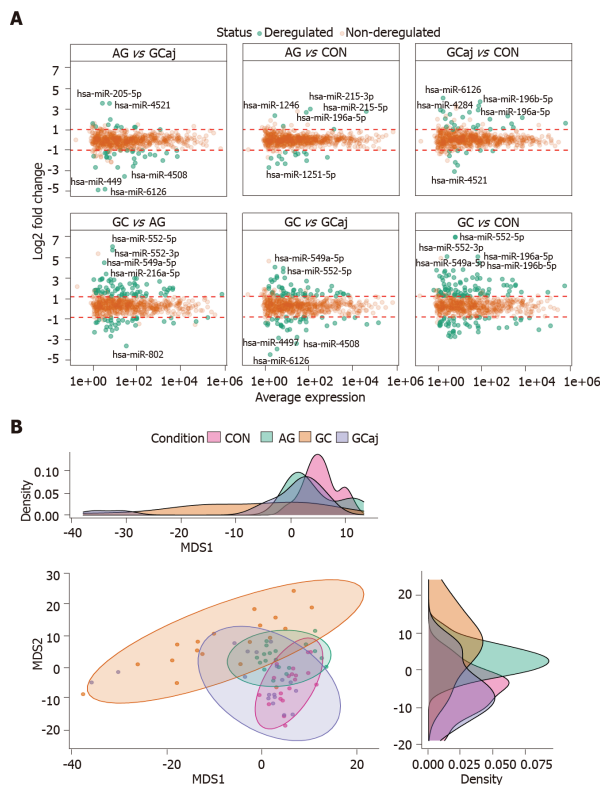
81

Figure 1 Results of microRNA differential expression analysis. A: Differentially expressed gastric tissue microRNAs among different conditions. *P*-adjusted < 0.05 and |log2 fold change| > 1; B: Multidimensional scaling plot based on normalized data showing a clustering corresponding to control, atrophic gastritis, gastric cancerous and adjacent tissues. The density plots show distributions of the first and second dimensions. CON: Control; AG: Atrophic gastritis; GC: Gastric cancerous; GCaj: Gastric adjacent tissue; MDS: Multidimensional scaling.

miRNAs), 18 (11.6%) - AG compared to GC, and 6 (3.9%) - AG compared to CON (Figure 2). Most of the deregulated miRNAs (*n* = 79, 68.7%) were similar between GC *vs* CON and GC *vs* AG comparison groups. 12 miRNAs (7.7%) were deregulated in both AG and GC groups when compared to CON. Four miRNAs (2.6%) were similarly deregulated between AG *vs* CON and AG *vs* GC groups. Finally, only 2 miRNAs (hsa-miR-129-1-3p and hsa-miR-196a-5p) (1.29%) were identified as deregulated between all comparison groups. AUC-ROC analysis revealed that expression level of hsa-miR-129-1-3p in tissues resulted in AUCs: 68.1%; 86.3%, and 78.1%, CON *vs* AG, CON *vs* GC, and AG *vs* GC, respectively (Figures 3A, 3B and 3C). In addition to this, expression level of hsa-miR-196a-5p could be significant for discrimination of CON *vs* AG, CON *vs* GC and AG *vs* GC and resulted in AUCs: 88.0%, 93.1% and 66.3% (Figures 3D, 3E and 3F).

Hsa-miR-129-1-3p and hsa-miR-196a-5p expression in the plasma follows the expression pattern of CON, AG, and GC tissues

Differential expression analysis of NGS data in tissue samples revealed that hsa-miR-129-1-3p was significantly down-regulated and hsa-miR-196a-5p was up-regulated in AG and GC tissues compared to CON ($P = 0.002$ and $P = 0.00018$; $P = 1.2 \times 10^{-5}$ and $P = 3.1 \times 10^{-5}$, respectively). Moreover, hsa-miR-129-1-3p was significantly down-regulated in the case of AG compared to GC ($P = 0.0021$) and reflected a stepwise process of a
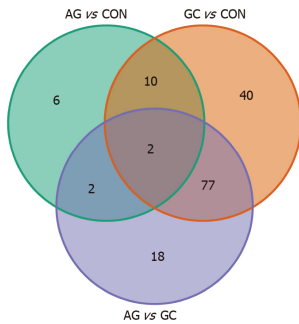
Figure 2 Venn diagram representing the number of commonly and uniquely differentially expressed microRNAs in three different comparison groups. *P*-adjusted < 0.05 and |log2 fold change| > 1. CON: Control; AG: Atrophic gastritis; GC: Gastric cancer.
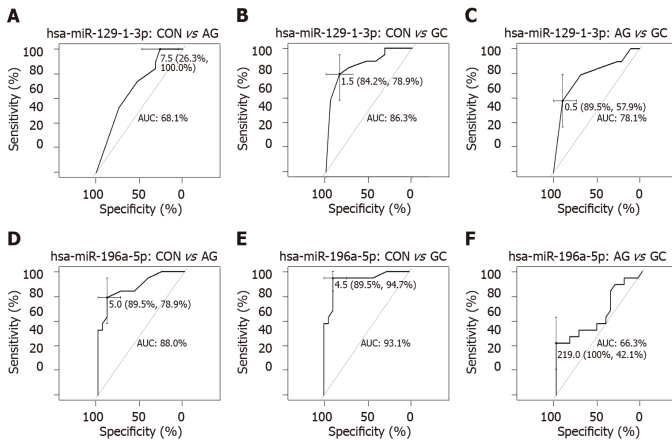

Figure 3 Receiver operating characteristic curves showing prediction performances of expression levels. A-C: Hsa-miR-129-1-3p; D-F: Hsa-miR-196a-5p in tissue samples between different comparison groups: Control *vs* atrophic gastritis; control *vs* gastric cancer; and atrophic gastritis *vs* gastric cancer. AUC: Area under the curve; CON: Control; AG: Atrophic gastritis; GC: Gastric cancer.

pathology (Figure 4A). Therefore, to identify whether the expression changes of these two miRNAs can be detected noninvasively in the body fluids of the patients, hsa-miR-129-1-3p and hsa-miR-196a-5p were selected for RT-qPCR analysis in plasma samples of independent cohort. The analysis showed similar expression patterns in the case of hsa-miR-129-1-3p, which was significantly down-regulated when comparing AG and GC groups (*P* = 0.024). There were no other significant findings between the groups (Figure 4B).

Hsa-miR-215-3p/5p and hsa-miR-934 may be associated with *H. pylori*-induced GC

To investigate role of miRNAs in AG atrophy progression (OLGA classification) and *H. pylori*-induced GC, differential miRNAs profile analysis in the subgroups of the study was performed. The analysis revealed a minor clustering in AG tissues corresponding to OLGA stages (Supplementary Figures 3A and 3H). *H. pylori* status in GC tissues (Supplementary Figure 3B). However, no significantly deregulated miRNAs were determined comparing I-II OLGA stages *vs* III-IV OLGA stages (AG tissue samples). On the other hand, analyzing GC group based on *H. pylori* infection status [*H. pylori* (neg.) *vs H. pylori* (pos.)], three miRNAs were shown to be significantly
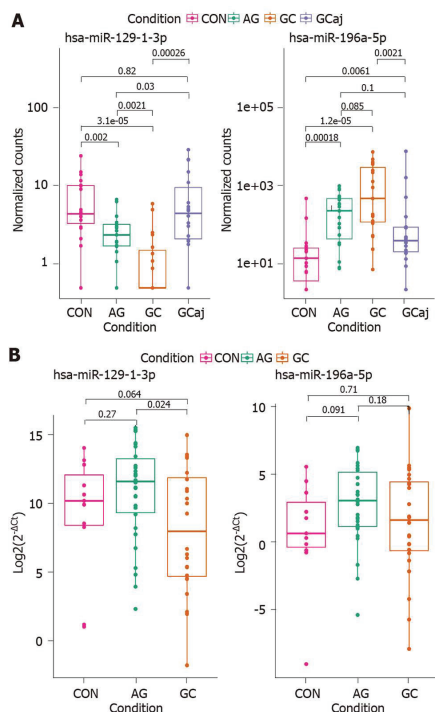
83

Figure 4 Hsa-miR-129-1-3p and hsa-miR-196a-5p expression levels in study comparison groups. A: Atrophic gastritis and gastric cancer tissue samples compared to controls; B: Atrophic gastritis and gastric cancer plasma samples compared to controls. Box plot graphs; boxes correspond to the median value and interquartile range. CON: Control; AG: Atrophic gastritis; GC: Gastric cancerous; GCaj: Gastric adjacent tissue.

deregulated: Hsa-miR-215-3p (log2FC = -4.52, *P*-adjusted = 0.02), hsa-miR-215-5p (log2FC = -4.00, *P*-adjusted = 0.02), and hsa-miR-934 (log2FC = 6.09, *P*-adjusted = 0.02).

DISCUSSION

This study represents comprehensive miRNome profiling of premalignant and malignant GC cases by implementing high throughput technologies such as NGS. Although there are several studies reporting profiles of GC tissue miRNAs[23,24], analysis of the association between miRNA expression and AG is very scarce reporting only individual miRNAs[25]. Moreover, based on small RNA-seq findings, two miRNAs were analyzed in subjects' plasma samples to investigate potential non-invasive markers. To our best knowledge this is the first study analyzing AG and GC tissue miRNomes in the subjects of European origin.

First, our study showed different profiles of deregulated miRNAs between tissue samples of studied groups. In total, 20 differentially expressed miRNAs were identified in AG and 129 - in GC comparing to CON; also 99 deregulated miRNAs - comparing GC and AG groups. MiRNAs such as hsa-miR-3131, hsa-miR-483, hsa-miR-150, hsa-miR-200a-3p, hsa-miR-873-5p were previously reported by the GC profiling studies of Pereira *et al*[23] and Assumpção *et al*[24]. Yet, we were able to identify number of novel miRNAs (of which hsa-miR-548ba, hsa-miR-4521, hsa-miR-549a were the most deregulated). There are no data showing the role of these novel miRNAs in inflammatory or tumorous processes of gastric tissue. However, recent studies have shown that hsa-miR-548ba was associated with bladder cancer, hsa-miR-549a with the

84

metastasis of renal cancer, and hsa-miR-4521 with *H. pylori* infection in esophageal epithelial cells[26-28]. Taking into consideration miRNome of AG, hsa-miR-3591-3p, hsa-miR-122-3p and hsa-miR-122-5p, hsa-miR-451a miRNAs were already reported by Liu *et al*[29], while the most deregulated miRNAs including hsa-miR-215, hsa-miR-4497, and hsa-miR-1251 were reported for the first time in our study. Previous research showed that hsa-miR-215-5p was deregulated in different lesions of the gastrointestinal tract (Barrett's esophagus, intraepithelial neoplastic lesions, ulcerative colitis)[30-32]. However, hsa-miR-4497 and hsa-miR-452 were not previously associated with AG but were reported to play an important role in GC development[33,34].

Next, we identified hsa-miR-215-3p and hsa-miR-215-5p to be down-regulated while hsa-miR-934 - up-regulated in GC group comparing negative and positive *H. pylori* infection status. Studies revealed the altered expression of various miRNAs in *H. pylori*-induced GC tissue samples, including miR-934, miR-146a, miR-375, miR-204[35-37]. Although, hsa-miR-215 deregulation was previously associated with GC[38-40], there is no data showing its link with *H. pylori* infection.

In addition to this, we showed that two miRNAs (hsa-miR-129-1-3p and hsa-miR-196a-5p) were gradually deregulated comparing all three study groups (CON, AG, and GC) which also corresponds to pathological cascade of GC. In concordance to our results, it has already been shown that hsa-miR-129-1-3p was down-regulated in GC tissues, function as a tumor suppressor in GC and even corresponds to the same expression pattern in gastric juice[41,42]. There is no data regarding the hsa-miR-196a expression in AG tissue, however, investigators have revealed that hsa-miR-196a is overexpressed in GC tissue, plasma, commercial cell lines and promotes cell proliferation[43,44]. ROC-AUC analysis suggests great potential of hsa-miR-196a-5p expression in tissue for discrimination of AG and GC in contrast to CON (AUC = 89.5% and AUC = 89.5%, respectively). Therefore, further studies are needed to confirm this finding.

Finally, selected miRNAs were analyzed in independent cohort of CON, AG, and GC plasma samples by using RT-qPCR. Results showed similar deregulation direction in plasma samples as in the tissue samples. However, significant differences were only determined comparing the expression of hsa-miR-129-1-3p between AG and GC suggesting its potential role in non-invasive diagnostics of malignant cases. No significant expression changes were observed between study groups and hsa-miR-196a-5p. Other studies have shown controversial results: Tsai *et al*[45] reported that miR-196a/b was up-regulated in both the plasma and tissue of metastatic GC patients, while miRNome profiling study revealed that miR-196a-5p was found to be down-regulated in plasma of patients with precursor lesions of GC compared to non-active gastritis[46].

In our study, using NGS and RT-qPCR techniques we have shown the distinct miRNome profiles of CON, AG, GC, GCaj tissues, and potential of specific miRNAs as non-invasive biomarkers. In addition to this, novel miRNAs not previously reported as AG or GC associated epigenetic markers were identified. We have shown that hsa-miR-196a-5p expression in tissue could be significant for discrimination between CON and AG or GC, confirmed hsa-miR-129-1-3p as non-invasive biomarker in disease progression monitoring, and showed that miRNAs could be a great candidate for future research of new diagnostic approaches.

CONCLUSION

In conclusion, we showed gradual deregulation of hsa-miR-196a-5p and hsa-miR-129-1-3p in the gastric carcinogenesis pathway and confirmed hsa-miR-129-1-3p as a possible non-invasive biomarker. We also found hsa-miR-215-3p/5p and hsa-miR-934 to be significantly deregulated in GC based on *H. pylori* infection status. These data provide novel insights into complex GC pathogenesis cascade which could be highly significant for future studies of new diagnostic GC targets.

ARTICLE HIGHLIGHTS

Research background

Gastric cancer (GC) is a complex disease arising from the interaction of environmental (*e.g.,* diet, smoking, *etc.*) and host-associated factors [*e.g., Helicobacter pylori* (*H. pylori*) infection, genetics, *etc.*]. Due to its silent course, it is also one of the most lethal cancers

85

worldwide as it is usually diagnosed at the advanced stages.

Research motivation

Novel biomarkers that would help to improve GC patients' diagnosis and prognosis are highly needed. Studies show that microRNAs (miRNAs) play an important role in many cancers and could be a promising biomarker or even therapeutic target.

Research objectives

The objectives of the study were to analyze whole miRNome profiles of control, premalignant and malignant gastric tissues, and select the potential miRNA markers that could have a potential for minimally invasive GC diagnostics.

Research methods

Total RNA from gastric tissue samples was subjected for small RNA sequencing (smRNA-seq). Plasma total circulating nucleic acids were used for the expression analysis of the most tissue deregulated miRNAs by real-time quantitative polymerase chain reaction. Statistical analysis involved the differential expression and discrimination analyses.

Research results

The abundance of altered expression miRNAs corresponded to a pathological cascade of GC development. Hsa-miR-129-1-3p and has-miR-196a-5p were shown to be deregulated in healthy-premalignant-malignant sequence. In addition to this, we showed that down-regulation of hsa-miR-129-1-3p could also be detected non-invasively in GC patients' plasma samples. Finally, results indicated that hsa-miR-215-3p/5p and hsa-miR-934 were significantly deregulated based on *H. pylori* infection status for GC patients.

Research conclusions

Gastric tissue miRNome study provides extensive profiling of control, premalignant and malignant cases. Based on smRNA-seq results several miRNAs were shown as potential gastric carcinogenesis (hsa-miR-196a-5p and hsa-miR-129-1-3p); and *H. Pylori*-related (hsa-miR-215-3p/5p and hsa-miR-934) biomarkers.

Research perspectives

This study provides novel insights into complex GC pathogenesis cascade and could serve as a reference for future research to support our findings.

REFERENCES

1   **Sung H**, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* 2021; **71**: 209-249 [PMID: 33538338 DOI: 10.3322/caac.21660]

2   **McLean MH**, El-Omar EM. Genetics of gastric cancer. *Nat Rev Gastroenterol Hepatol* 2014; **11**: 664-674 [PMID: 25134511 DOI: 10.1038/nrgastro.2014.143]

3   **Tan P**, Yeoh KG. Genetics and Molecular Pathogenesis of Gastric Adenocarcinoma. *Gastroenterology* 2015; **149**: 1153-1162.e3 [PMID: 26073375 DOI: 10.1053/j.gastro.2015.05.059]

4   **Bartel DP**. Metazoan MicroRNAs. *Cell* 2018; **173**: 20-51 [PMID: 29570994 DOI: 10.1016/j.cell.2018.03.006]

5   **Bartel DP**. MicroRNAs: target recognition and regulatory functions. *Cell* 2009; **136**: 215-233 [PMID: 19167326 DOI: 10.1016/j.cell.2009.01.002]

6   **Bracken CP**, Scott HS, Goodall GJ. A network-biology perspective of microRNA function and dysfunction in cancer. *Nat Rev Genet* 2016; **17**: 719-732 [PMID: 27795564 DOI: 10.1038/nrg.2016.134]

7   **Farazi TA**, Hoell JI, Morozov P, Tuschl T. MicroRNAs in human cancer. *Adv Exp Med Biol* 2013; **774**: 1-20 [PMID: 23377965 DOI: 10.1007/978-94-007-5590-1_1]

8   **Gyvyte U**, Juzenas S, Salteniene V, Kupcinskas J, Poskiene L, Kucinskas L, Jarmalaite S, Stuopelyte K, Steponaitiene R, Hemmrich-Stanisak G, Hübenthal M, Link A, Franke S, Franke A, Pangonyte D, Lesauskaite V, Kupcinskas L, Skieceviciene J. MiRNA profiling of gastrointestinal stromal tumors by next-generation sequencing. *Oncotarget* 2017; **8**: 37225-37238 [PMID: 28402935 DOI: 10.18632/oncotarget.16664]

9   **Streleckiene G**, Inciuraite R, Juzenas S, Salteniene V, Steponaitiene R, Gyvyte U, Kiudelis G, Leja M, Ruzgys P, Satkauskas S, Kupcinskiene E, Franke S, Thon C, Link A, Kupcinskas J, Skieceviciene J. miR-20b and miR-451a Are Involved in Gastric Carcinogenesis through the PI3K/AKT/mTOR

Signaling Pathway: Data from Gastric Cancer Patients, Cell Lines and Ins-Gas Mouse Model. *Int J Mol Sci* 2020; **21** [PMID: 32013265 DOI: 10.3390/ijms21030877]

10   **Link A**, Kupcinskas J. MicroRNAs as non-invasive diagnostic biomarkers for gastric cancer: Current insights and future perspectives. *World J Gastroenterol* 2018; **24**: 3313-3329 [PMID: 30122873 DOI: 10.3748/wjg.v24.i30.3313]

11   **Juzėnas S**, Saltenienė V, Kupcinskas J, Link A, Kiudelis G, Jonaitis L, Jarmalaite S, Kupcinskas L, Malfertheiner P, Skieceviciene J. Analysis of Deregulated microRNAs and Their Target Genes in Gastric Cancer. *PLoS One* 2015; **10**: e0132327 [PMID: 26172537 DOI: 10.1371/journal.pone.0132327]

12   **Rugge M**, Correa P, Di Mario F, El-Omar E, Fiocca R, Geboes K, Genta RM, Graham DY, Hattori T, Malfertheiner P, Nakajima S, Sipponen P, Sung J, Weinstein W, Vieth M. OLGA staging for gastritis: a tutorial. *Dig Liver Dis* 2008; **40**: 650-658 [PMID: 18424244 DOI: 10.1016/j.dld.2008.02.030]

13   **Amin MB**, Greene FL, Edge SB, Compton CC, Gershenwald JE, Brookland RK, Meyer L, Gress DM, Byrd DR, Winchester DP. The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging. *CA Cancer J Clin* 2017; **67**: 93-99 [PMID: 28094848 DOI: 10.3322/caac.21388]

14   **Lauren P**. The two histological main types of gastric carcinoma: diffuse and so-called intestinal-type carcinoma. An attempt at a histo-clinical classification. *Acta Pathol Microbiol Scand* 1965; **64**: 31-49 [PMID: 14320675 DOI: 10.1111/apm.1965.64.1.31]

15   **Kozomara A**, Birgaoanu M, Griffiths-Jones S. miRBase: from microRNA sequences to function. *Nucleic Acids Res* 2019; **47**: D155-D162 [PMID: 30423142 DOI: 10.1093/nar/gky1141]

16   **Di Tommaso P**, Chatzou M, Floden EW, Barja PP, Palumbo E, Notredame C. Nextflow enables reproducible computational workflows. *Nat Biotechnol* 2017; **35**: 316-319 [PMID: 28398311 DOI: 10.1038/nbt.3820]

17   **Langmead B**, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 2009; **10**: R25 [PMID: 19261174 DOI: 10.1186/gb-2009-10-3-r25]

18   **Ewels P**, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 2016; **32**: 3047-3048 [PMID: 27312411 DOI: 10.1093/bioinformatics/btw354]

19   **Robinson MD**, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010; **26**: 139-140 [PMID: 19910308 DOI: 10.1093/bioinformatics/btp616]

20   **Love MI**, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014; **15**: 550 [PMID: 25516281 DOI: 10.1186/s13059-014-0550-8]

21   **Ranstam J**. Multiple P-values and Bonferroni correction. *Osteoarthritis Cartilage* 2016; **24**: 763-764 [PMID: 26802548 DOI: 10.1016/j.joca.2016.01.008]

22   **Livak KJ**, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 2001; **25**: 402-408 [PMID: 11846609 DOI: 10.1006/meth.2001.1262]

23   **Pereira A**, Moreira F, Vinasco-Sandoval T, Cunha A, Vidal A, Ribeiro-Dos-Santos AM, Pinto P, Magalhães L, Assumpção M, Demachki S, Santos S, Assumpção P, Ribeiro-Dos-Santos Â. miRNome Reveals New Insights Into the Molecular Biology of Field Cancerization in Gastric Cancer. *Front Genet* 2019; **10**: 592 [PMID: 31275362 DOI: 10.3389/fgene.2019.00592]

24   **Assumpção MB**, Moreira FC, Hamoy IG, Magalhães L, Vidal A, Pereira A, Burbano R, Khayat A, Silva A, Santos S, Demachki S, Ribeiro-Dos-Santos Â, Assumpção P. High-Throughput miRNA Sequencing Reveals a Field Effect in Gastric Cancer and Suggests an Epigenetic Network Mechanism. *Bioinform Biol Insights* 2015; **9**: 111-117 [PMID: 26244015 DOI: 10.4137/BBI.S24066]

25   **Link A**, Schirrmeister W, Langner C, Varbanova M, Bornschein J, Wex T, Malfertheiner P. Differential expression of microRNAs in preneoplastic gastric mucosa. *Sci Rep* 2015; **5**: 8270 [PMID: 25652892 DOI: 10.1038/srep08270]

26   **Zhao F**, Ge YZ, Zhou LH, Xu LW, Xu Z, Ping WW, Wang M, Zhou CC, Wu R, Jia RP. Identification of hub miRNA biomarkers for bladder cancer by weighted gene coexpression network analysis. *Onco Targets Ther* 2017; **10**: 5551-5559 [PMID: 29200870 DOI: 10.2147/OTT.S146479]

27   **Xuan Z**, Chen C, Tang W, Ye S, Zheng J, Zhao Y, Shi Z, Zhang L, Sun H, Shao C. TKI-Resistant Renal Cancer Secretes Low-Level Exosomal miR-549a to Induce Vascular Permeability and Angiogenesis to Promote Tumor Metastasis. *Front Cell Dev Biol* 2021; **9**: 689947 [PMID: 34179017 DOI: 10.3389/fcell.2021.689947]

28   **Teng G**, Dai Y, Chu Y, Li J, Zhang H, Wu T, Shuai X, Wang W. Helicobacter pylori induces caudal-type homeobox protein 2 and cyclooxygenase 2 expression by modulating microRNAs in esophageal epithelial cells. *Cancer Sci* 2018; **109**: 297-307 [PMID: 29215765 DOI: 10.1111/cas.13462]

29   **Liu H**, Li PW, Yang WQ, Mi H, Pan JL, Huang YC, Hou ZK, Hou QK, Luo Q, Liu FB. Identification of non-invasive biomarkers for chronic atrophic gastritis from serum exosomal microRNAs. *BMC Cancer* 2019; **19**: 129 [PMID: 30736753 DOI: 10.1186/s12885-019-5328-7]

30   **Fassan M**, Croce CM, Rugge M. miRNAs in precancerous lesions of the gastrointestinal tract. *World J Gastroenterol* 2011; **17**: 5231-5239 [PMID: 22219591 DOI: 10.3748/wjg.v17.i48.5231]

31   **Fassan M**, Volinia S, Palatini J, Pizzi M, Baffa R, De Bernard M, Battaglia G, Parente P, Croce CM, Zaninotto G, Ancona E, Rugge M. MicroRNA expression profiling in human Barrett's carcinogenesis. *Int J Cancer* 2011; **129**: 1661-1670 [PMID: 21128279 DOI: 10.1002/ijc.25823]

87

32   **Wijnhoven BP**, Hussey DJ, Watson DI, Tsykin A, Smith CM, Michael MZ; South Australian Oesophageal Research Group. MicroRNA profiling of Barrett's oesophagus and oesophageal adenocarcinoma. *Br J Surg* 2010; **97**: 853-861 [PMID: 20301167 DOI: 10.1002/bjs.7000]

33   **Bibi F**, Naseer MI, Alvi SA, Yasir M, Jiman-Fatani AA, Sawan A, Abuzenadah AM, Al-Qahtani MH, Azhar EI. microRNA analysis of gastric cancer patients from Saudi Arabian population. *BMC Genomics* 2016; **17**: 751 [PMID: 27766962 DOI: 10.1186/s12864-016-3090-7]

34   **Yin C**, Zheng X, Xiang H, Li H, Gao M, Meng X, Yang K. Differential expression profile analysis of cisplatinregulated miRNAs in a human gastric cancer cell line. *Mol Med Rep* 2019; **20**: 1966-1976 [PMID: 31257509 DOI: 10.3892/mmr.2019.10430]

35   **Chang H**, Kim N, Park JH, Nam RH, Choi YJ, Lee HS, Yoon H, Shin CM, Park YS, Kim JM, Lee DH. Different microRNA expression levels in gastric cancer depending on Helicobacter pylori infection. *Gut Liver* 2015; **9**: 188-196 [PMID: 25167801 DOI: 10.5009/gnl13371]

36   **Prinz C**, Mese K, Weber D. MicroRNA Changes in Gastric Carcinogenesis: Differential Dysregulation during *Helicobacter pylori* and EBV Infection. *Genes (Basel)* 2021; **12** [PMID: 33921696 DOI: 10.3390/genes12040597]

37   **Wang F**, Sun G, Zou Y, Zhong F, Ma T, Li X. Protective role of Helicobacter pylori infection in prognosis of gastric cancer: evidence from 2,454 patients with gastric cancer. *PLoS One* 2013; **8**: e62440 [PMID: 23667477 DOI: 10.1371/journal.pone.0062440]

38   **Jin Z**, Selaru FM, Cheng Y, Kan T, Agarwal R, Mori Y, Olaru AV, Yang J, David S, Hamilton JP, Abraham JM, Harmon J, Duncan M, Montgomery EA, Meltzer SJ. MicroRNA-192 and -215 are upregulated in human gastric cancer *in vivo* and suppress ALCAM expression in vitro. *Oncogene* 2011; **30**: 1577-1585 [PMID: 21119604 DOI: 10.1038/onc.2010.534]

39   **Deng Y**, Huang Z, Xu Y, Jin J, Zhuo W, Zhang C, Zhang X, Shen M, Yan X, Wang L, Wang X, Kang Y, Si J, Zhou T. MiR-215 modulates gastric cancer cell proliferation by targeting RB1. *Cancer Lett* 2014; **342**: 27-35 [PMID: 23981575 DOI: 10.1016/j.canlet.2013.08.033]

40   **Li N**, Zhang QY, Zou JL, Li ZW, Tian TT, Dong B, Liu XJ, Ge S, Zhu Y, Gao J, Shen L. miR-215 promotes malignant progression of gastric cancer by targeting RUNX1. *Oncotarget* 2016; **7**: 4817-4828 [PMID: 26716895 DOI: 10.18632/oncotarget.6736]

41   **Wang D**, Luo L, Guo J. miR-129-1-3p inhibits cell migration by targeting BDKRB2 in gastric cancer. *Med Oncol* 2014; **31**: 98 [PMID: 25008064 DOI: 10.1007/s12032-014-0098-1]

42   **Yu X**, Luo L, Wu Y, Yu X, Liu Y, Zhao X, Zhang X, Cui L, Ye G, Le Y, Guo J. Gastric juice miR-129 as a potential biomarker for screening gastric cancer. *Med Oncol* 2013; **30**: 365 [PMID: 23307240 DOI: 10.1007/s12032-012-0365-y]

43   **Treece AL**, Duncan DL, Tang W, Elmore S, Morgan DR, Dominguez RL, Speck O, Meyers MO, Gulley ML. Gastric adenocarcinoma microRNA profiles in fixed tissue and in plasma reveal cancer-associated and Epstein-Barr virus-related expression patterns. *Lab Invest* 2016; **96**: 661-671 [PMID: 26950485 DOI: 10.1038/labinvest.2016.33]

44   **Sun M**, Liu XH, Li JH, Yang JS, Zhang EB, Yin DD, Liu ZL, Zhou J, Ding Y, Li SQ, Wang ZX, Cao XF, De W. MiR-196a is upregulated in gastric cancer and promotes cell proliferation by downregulating p27(kip1). *Mol Cancer Ther* 2012; **11**: 842-852 [PMID: 22343731 DOI: 10.1158/1535-7163.MCT-11-1015]

45   **Tsai MM**, Wang CS, Tsai CY, Huang CG, Lee KF, Huang HW, Lin YH, Chi HC, Kuo LM, Lu PH, Lin KH. Circulating microRNA-196a/b are novel biomarkers associated with metastatic gastric cancer. *Eur J Cancer* 2016; **64**: 137-148 [PMID: 27420607 DOI: 10.1016/j.ejca.2016.05.007]

46   **Lario S**, Brunet-Vega A, Quílez ME, Ramírez-Lázaro MJ, Lozano JJ, García-Martínez L, Pericay C, Miquel M, Junquera F, Campo R, Calvet X. Expression profile of circulating microRNAs in the Correa pathway of progression to gastric cancer. *United European Gastroenterol J* 2018; **6**: 691-701 [PMID: 30083331 DOI: 10.1177/2050640618759433]

88

A2

Title: miR-20b and miR-451a are involved in gastric carcinogenesis through the PI3K/AKT/mTOR signaling pathway: data from gastric cancer patients, cell lines and ins-gas mouse model

Authors: Streleckienė, Greta; Inčiūraitė, Rūta; Juzėnas, Simonas; Šaltenienė, Violeta; Gyvytė, Ugnė; Kiudelis, Gediminas; Leja, Mārcis; Ruzgys, Paulius; Šatkauskas, Saulius; Kupčinskienė, Eugenija; Franke, Sabine; Thon, Cosima; Link, Alexander; Kupčinskas, Juozas; Skiecevičienė, Jurgita

International journal of molecular sciences 21, 3 (2020)

Article

miR-20b and miR-451a Are Involved in Gastric Carcinogenesis through the PI3K/AKT/mTOR Signaling Pathway: Data from Gastric Cancer Patients, Cell Lines and Ins-Gas Mouse Model

Greta Streleckiene [1], Ruta Inciuraite [1], Simonas Juzenas [1,2], Violeta Salteniene [1],
Ruta Steponaitiene [1], Ugne Gyvyte [1], Gediminas Kiudelis [1], Marcis Leja [3,4,5,6],
Paulius Ruzgys [7], Saulius Satkauskas [7], Eugenija Kupcinskiene [8], Sabine Franke [9],
Cosima Thon [10], Alexander Link [10], Juozas Kupcinskas [1,11,†] and Jurgita Skieceviciene [1,*,†]

[1] Institute for Digestive Research, Academy of Medicine, Lithuanian University of Health Sciences,
     Kaunas LT-50161, Lithuania; greta.streleckiene@lsmuni.lt (G.S.); ruta.inciuraite@lsmuni.lt (R.I.);
     simonas.juzenas@lsmuni.lt (S.J.); violeta.salteniene@lsmuni.lt (V.S.); ruta.steponaitiene@lsmuni.lt (R.S.);
     ugne.gyvyte@lsmuni.lt (U.G.); gediminaskiudelis@gmail.com (G.K.); j_kupcinskas@yahoo.com (J.K.)
[2] Institute of Clinical Molecular Biology, Christian-Albrechts-University of Kiel, 24105 Kiel, Germany
[3] Institute for Clinical and Preventive Medicine, University of Latvia, Riga LV-1586, Latvia; marcis.leja@lu.lv
[4] Faculty of Medicine, University of Latvia, Riga LV-1586, Latvia
[5] Department of Research, Riga East University Hospital, Riga LV-1038, Latvia
[6] Digestive Diseases Centre GASTRO, Riga LV-1079, Latvia
[7] Biophysical Research Group, Faculty of Natural Sciences, Vytautas Magnus University,
     Kaunas LT-44404, Lithuania; paulius.ruzgys@vdu.lt (P.R.); saulius.satkauskas@vdu.lt (S.S.)
[8] Department of Biology, Faculty of Nature Sciences, Vytautas Magnus University,
     Kaunas LT-44404, Lithuania; eugenija.kupcinskiene@vdu.lt
[9] Institute of Pathology, Otto-von-Guericke University, 39120 Magdeburg, Germany;
     sabine.franke@med.ovgu.de
[10] Department of Gastroenterology, Hepatology and Infectious Diseases, Otto-von-Guericke University,
     39120 Magdeburg, Germany; cosima.thon@med.ovgu.de (C.T.); alexander.link@med.ovgu.de (A.L.)
[11] Department of Gastroenterology, Academy of Medicine, Lithuanian University of Health Sciences,
     Kaunas LT-50161, Lithuania
*   Correspondence: jurgita.skieceviciene@lsmuni.lt; Tel.: +370-37-327236
†   These authors contributed equally to this work.

check for
updates

Abstract: Gastric cancer (GC) is one of the most common and lethal gastrointestinal malignancies worldwide. Many studies have shown that development of GC and other malignancies is mainly driven by alterations of cellular signaling pathways. MicroRNAs (miRNAs) are small noncoding molecules that function as tumor-suppressors or oncogenes, playing an essential role in a variety of fundamental biological processes. In order to understand the functional relevance of miRNA dysregulation, studies analyzing their target genes are of major importance. Here, we chose to analyze two miRNAs, miR-20b and miR-451a, shown to be deregulated in many different malignancies, including GC. Deregulated expression of miR-20b and miR-451a was determined in GC cell lines and the INS-GAS mouse model. Using Western Blot and luciferase reporter assay we determined that miR-20b directly regulates expression of *PTEN* and *TXNIP*, and miR-451a: *CAV1* and *TSC1*. Loss-of-function experiments revealed that down-regulation of miR-20b and up-regulation of miR-451a expression exhibits an anti-tumor effect in vitro (miR-20b: reduced viability, colony formation, increased apoptosis rate, and miR-451a: reduced colony forming ability). To summarize, the present study identified that expression of miR-20b and miR-451a are deregulated in vitro and in vivo and have a tumor suppressive role in GC through regulation of the PI3K/AKT/mTOR signaling pathway.

1. Introduction

Gastric cancer (GC) is one of the most prevalent malignancies and a leading cause of cancer-related mortality worldwide [1]. The development and progression of GC is a multistep process involving accumulation of genetic mutations and alterations in proto-oncogenes or tumor-suppressor genes [2]. One of the major issues in clinical settings of GC is a poor survival rate and prognosis of patients, as the tumor is diagnosed at late stages of the disease. Therefore, further insight into the molecular mechanisms underlying GC progression may help to identify novel therapeutic targets and improve the prognosis of GC.

MicroRNAs (miRNAs) are small noncoding molecules that, being part of a so-called RNA-induced silencing complex (RISC), post-transcriptionally regulate gene expression [3]. Dependent on the target gene and background conditions, miRNAs may function as either tumor-suppressors which suppress protein-coding oncogenes, or as oncogenic miRNAs which negatively regulate known tumor-suppressor genes [4]. In this way they play an essential role in a variety of fundamental biological and pathological processes [3]. Moreover, miRNA-based therapy is being tested as a potential strategy in cancer treatment [5–8]. A number of GC studies have reported specific signatures of deregulated miRNAs belonging to miR-17, miR-19, miR-21, miR-223, miR-135, and other families, and their diagnostic and prognostic potential [9–12]. However, in order to understand functional relevance of miRNA dysregulation, studies analyzing their target genes are of major importance.

Here, we chose to analyze two miRNAs, miR-20b and miR-451a, shown to be deregulated in many different malignancies, including GC [13–17]. Previous functional studies have revealed that these miRNAs may exert their biological role through mediating tumor formation, maintenance, and metastasis [18–21]. However, the role of these miRNAs and possible target genes in GC remain poorly investigated. In this study, we found that inhibition of miR-20b and overexpression of miR-451a had a tumor-suppressive role in GC through the regulation of genes involved in the PI3K/AKT/mTOR signaling pathway. Inhibition of miR-20b reduced GC cell viability, proliferation and promoted early cell apoptosis; whereas overexpression of miR-451a reduced GC cell growth. These findings reveal the important roles of miR-20b and miR-451a in GC progression, which may be used to develop a beneficial strategy for future cancer therapy.

2. Results

2.1. Aberrant Expression of miR-20b and miR-451a In Vivo

First, the expression level of miR-20b and miR-451a was determined *in vivo.* For this purpose, biopsy samples of GC patients, healthy controls, and INS-GAS mice samples were analyzed.

MiRNA expression analysis in GC samples showed a significant increase of miR-20b ($p = 0.026$) and decrease of miR-451a ($p = 0.039$) expression compared to controls (Figure 1).

Moreover, GC analysis of INS-GAS mice showed a gender specific miR-20b expression pattern following *H. pylori* infection. Only male mice showed significantly higher miR-20b expression for all time points ($p = 0.029$). There was a stepwise increase in miR-20b expression during the different time points from 12 to 50 weeks with the highest difference at 50 weeks ($p = 0.003$) (Figure 2B). No significant changes in miR-451a expression were observed.
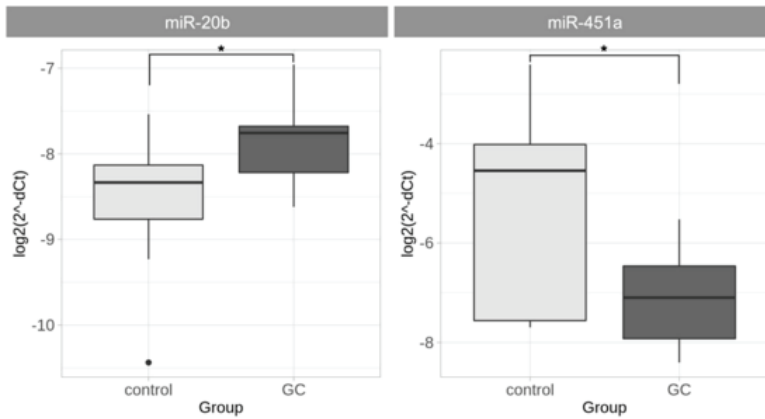
Figure 1. miR-20b and miR-451 expression analysis in GC and control tissues (normalized delta Ct (dCt) values are presented in logarithmic scale). miR-20b expression was significantly increased and miR-451a gene expression was significantly decreased in GC tissue compared to control tissue (* $p < 0.05$).
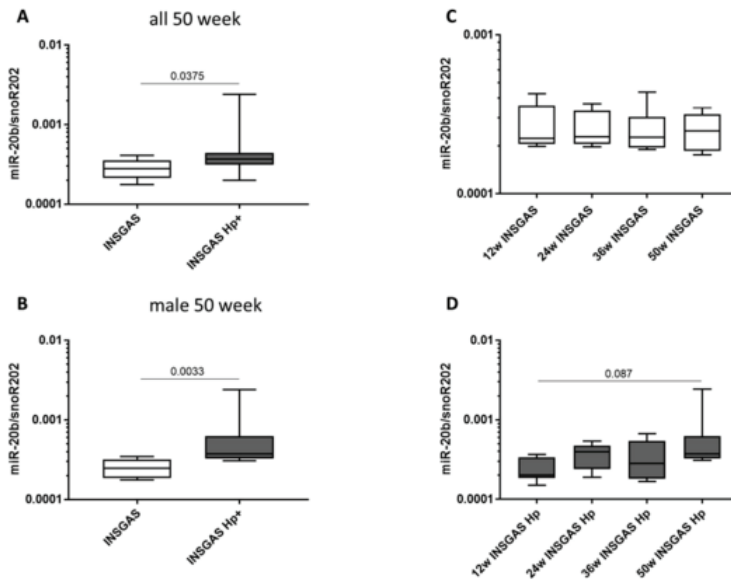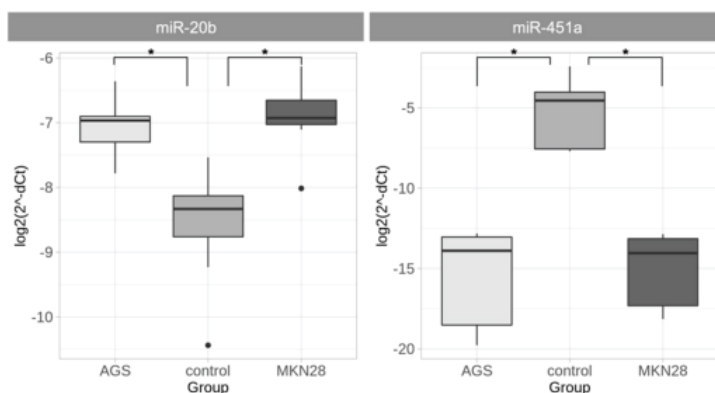

Figure 2. miR-20b expression analysis in vivo using INS-GAS mouse model. (**A**) MiR-20b expression level in male and female mice comparing INS-GAS and *H. pylori* infected INS-GAS mice at 50 weeks. Analysis showed significant increase in miR-20b expression ($p = 0.0375$). (**B**) MiR-20b expression level in male mice comparing INS-GAS and *H. pylori* infected INS-GAS mice at 50 weeks. Results revealed significant increase in miR-20b expression and gender specific expression pattern ($p = 0.033$). (**C** and **D**) MiR-20b expression at 12, 24, 36 and 50 weeks. A stepwise increase in miR-20b expression was determine during the different time points with highest difference at 50 weeks ($p = 0.003$).

2.2. Aberrant Expression of miR-20b and miR-451a in Vitro

The expression level of miR-20b and miR-451a was also determined in vitro in AGS and MKN28 cell lines compared to normal gastric tissue. Expression level of miR-20b was significantly up-regulated ($p = 6.35 \times 10^{-5}$ and $p = 1.03 \times 10^{-4}$, AGS and MKN28, respectively) and miR-451a was down-regulated ($p = 2.58 \times 10^{-5}$ and $p = 2.58 \times 10^{-5}$, AGS and MKN28, respectively) in both GC cell lines compared with control gastric tissue (Figure 3). There was no difference on microRNA expression patterns between the two cell lines.

Figure 3. miR-20b and miR-451 expression analysis in AGS and MKN28 cell lines (normalized delta Ct (dCt) values are presented in logarithmic scale). miR-20b expression was significantly increased in both gastric cancer cell cultures compared to control tissue group and miR-451a gene expression was significantly decreased in both gastric cancer cell cultures AGS and MKN28 compared to control tissue group (* $p < 0.05$).

These data suggest that miR-20b may potentially act as onco-miRNA and target tumour-suppressor genes while miR-451a acts as tumour-suppressor miRNA and target proto-oncogenes. Based on the observed expression patterns, mimic of miR-451a and inhibitor of miR-20b were chosen for loss-of-function study for both cell cultures.

2.3. Inhibition of miR-20b Reduced Cell Viability and Proliferation

To characterize functional importance of tested miRNAs viability and proliferation of AGS and MKN28 cell lines were tested using MTT assay, 48 h and 72 h after transfection. Reduced cell viability (by 22.1%) was observed in the AGS cell line 72 h after transfection with anti-miR-20b ($p = 0.029$) (Figure 4), whereas no significant changes were observed in MKN28 cell line. Overexpression of miR-451a had no significant effect on viability and proliferation in AGS or MKN28 cells.

2.4. Inhibition of miR-20b and Overexpression of miR-451a Dramatically Reduced Colony Formation Rate

To examine the role of miR-20b and miR-451a in GC cell growth, colony formation assay was conducted. The number of colonies reduced dramatically (AGS by 73.8%; $p = 2 \times 10^{-4}$ and MKN28 by 60.1%; $p = 0.021$) after inhibition of miR-20b compared to cells transfected with miRNA control (Figure 4). Transfection of GC cells with miR-451a mimic significantly reduced the number of colonies by 50.6% in MKN28 cell culture ($p = 0.043$) (Figure 5).
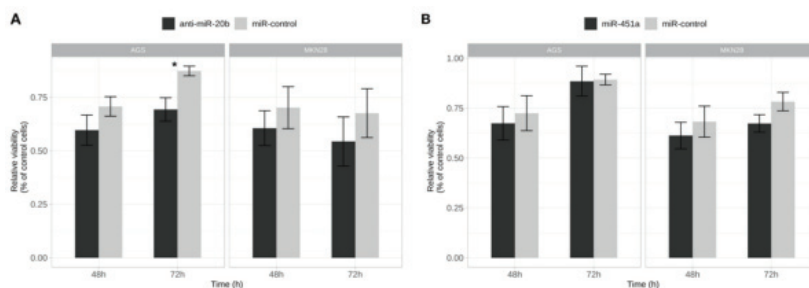
Figure 4. Exogenously altered expression of miR-20b effected cell viability. (**A**) Relative viability changes of AGS and MKN28 cells after inhibition of miR-20b 48 h and 72 h after transfection. Significant results evaluated in AGS cell lines 72h after transfection (* $p < 0.05$); (**B**) Relative viability changes of AGS and MKN28 cells after increased expression of miR-451a 48 h and 72 h after transfection. Data from four independent experiments.
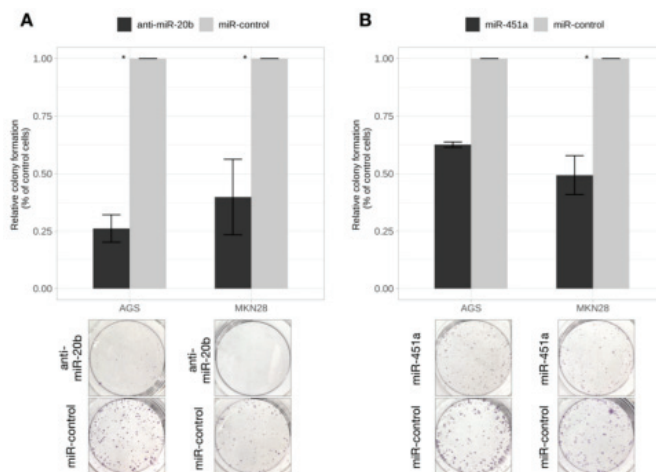

Figure 5. Exogenously altered expression of miR-20b and miR-451a affected colony formation. (**A**) Relative colony formation (of control cells) in AGS and MKN28 cells after inhibition of miR-20b. Formation of colonies reduced significantly in both AGS and MKN28 cell cultures compared to miR-control (* $p < 0.05$); (**B**) Relative colony formation (of control cells) in AGS and MKN28 cells after increased expression of miR-451a. Formation of colonies reduced significantly in both AGS and MKN28 cell cultures compared to miR-control (* $p < 0.05$). Representative pictures of fixed colonies are presented at the bottom of a figure. Data from five independent experiments.

2.5. Inhibition of miR-20b Increased Cell Apoptosis Rate

In order to investigate changes in early apoptosis and cell death rates flow cytometry-based apoptosis assay was employed, where annexin V-FITC positive cells were considered early apoptotic and annexin V-FITC/ PI positive as necrotic cells. A slightly increased rate of apoptotic cells (by 12.5%) was detected in AGS cell culture ($p = 0.040$) after inhibition of miR-20b (Figure 6). However, a decrease in live cells did not reach statistical significance. No effect on live, apoptotic, or necrotic populations was observed in MKN28 cell line after transfection with anti-miR-20b and in both cell cultures after transfection with miR-451a mimic.
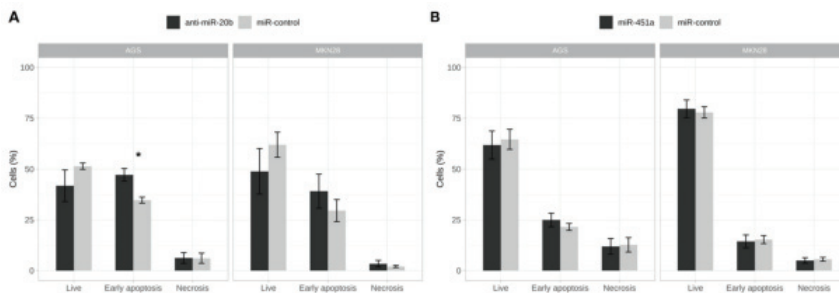
Figure 6. Exogenously altered expression of miR-20b effected proportion of cells undergoing process of early apoptosis. (**A**) Changes of apoptosis rates in AGS and MKN28 cell cultures after inhibition of miR-20b. Proportion of early apoptotic cells increased statistically significantly in AGS cells line (* $p < 0.05$); (**B**) Changes of apoptosis rates in AGS and MKN28 cell cultures after increased expression of miR-451a. Data from three independent experiments.

2.6. miR-20b and miR-451a Have No Effect In Vitro On Cell Migration Processes

The effect of miR-20b and miR-451a on the migration of GC cells was analyzed using wound healing assay. However, no significant changes in migration were determined in both GC cell lines after transfection with anti-miR-20b and miR-451a.

2.7. PTEN and TXNIP Are Direct Targets of miR-20b

IRF1, *PTEN*, and *TXNIP* were selected in silico as potential target genes of miR-20b. Protein expression analysis revealed that inhibition of miR-20b resulted in a significant increase of PTEN expression 72 h after transfection ($p = 0.011$) in AGS cell line and increase of TXNIP protein expression 72 h after transfection ($p = 0.025$) in MKN28 cells (Figure 7); whereas, no effect on IRF1 protein level was observed in both cell lines.
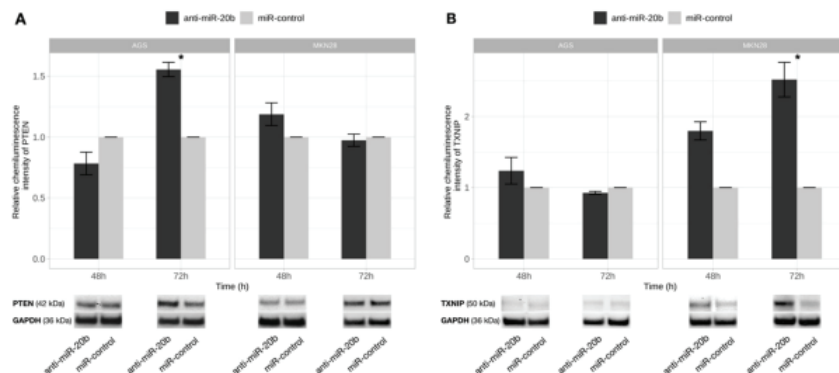
Figure 7. PTEN and TXNIP proteins level changes after exogenous miR-20b expression inhibition. PTEN (**A**) and TXNIP (**B**) protein expression comparison 48 h and 72 h after transfection in AGS and MKN28 cell cultures transfected with anti-miR-20b and miR-control. Significant PTEN protein level increase was determined 72 h after transfection in AGS cell culture and TXNIP- 72 h after transfection in MKN28 cell culture. Representative pictures of PTEN and TXNIP proteins signals detected by Western Blot presented at the bottom of a figure (* $p < 0.05$). Data from four independent experiments.

To validate the binding specificity of miR-20b to *PTEN* and *TXNIP*, a luciferase reporter system containing 3′ UTR-wt and 3′ UTR-mut regions of the genes was used. AGS cells were co-transfected with reporter vector and miR-20b mimic or negative mimic control. As a result, miR-20b clearly suppressed luciferase activity both in *PTEN*-3′ UTR-wt and *TXNIP*-3′UTR-wt compared to the control ($p = 0.006$ and $p = 4 \times 10^{-4}$, respectively), whereas firefly luciferase activity did not change for both mut-type vectors (Figure 8).
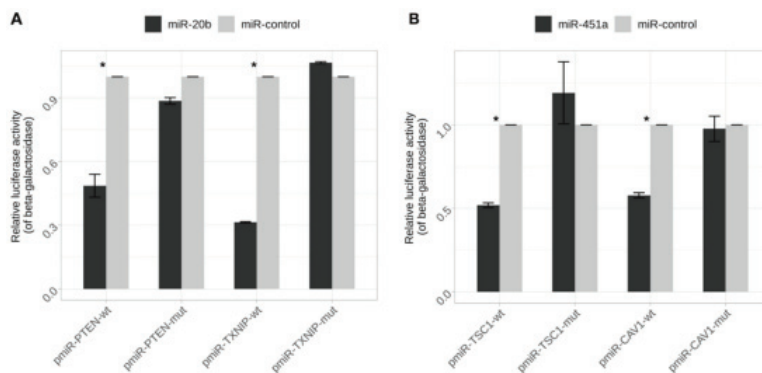
Figure 8. Estimation of direct interaction of investigated miRNAs and predicted target genes 3′ UTR by luciferase reporter assay. (**A**) AGS cell line was cotransfected with miR-20b (or miR-control) and pmiR-PTEN-wt/mut or pmiR-TXNIP-wt/mut vectors. Significant signal decrease was determined in cells transfected with either PTEN or TXNIP wt vectors (* $p < 0.05$); (**B**) AGS cell line was cotransfected with miR-451a (or miR-control) and pmiR-CAV1-wt/mut or pmiR-TSC1-wt/mut vectors. Significant signal decrease was determined in cells transfected with either CAV1 or TSC1 wt vectors (* $p < 0.05$). Luciferase activity was normalized by the beta-galactosidase signals. Results are shown as fold change relative to the negative control. Data from three independent experiments.

2.8. miR-451a Directly Regulate the Expression of CAV1 and TSC1

Using in silico prediction tools as downstream targets of miR-451a, *CAV1* and *TSC1* were selected. Upregulation of miR-451a reduced CAV1 and TSC1 proteins expression in AGS cell culture 72 h after transfection ($p = 0.011$ and $p = 0.024$, respectively) (Figure 9). No significant changes in protein level of target genes was observed in MKN28 cell line.

Direct binding specificity of miR-451a to *CAV1* and *TSC1* was evaluated using luciferase reporter system containing 3′ UTR-wt and 3′ UTR-mut regions. AGS cells were co-transfected with miR-451a mimic or negative mimic control and reporter vector. The results indicated that miR-451a significantly reduced firefly luciferase activity both in *CAV1*-3′ UTR-wt and *TSC1*-3′UTR-wt ($p = 0.007$ and $p = 0.009$, respectively) compared to the control. Firefly luciferase activity did not change in cells transfected with the mut-type vectors (Figure 8).
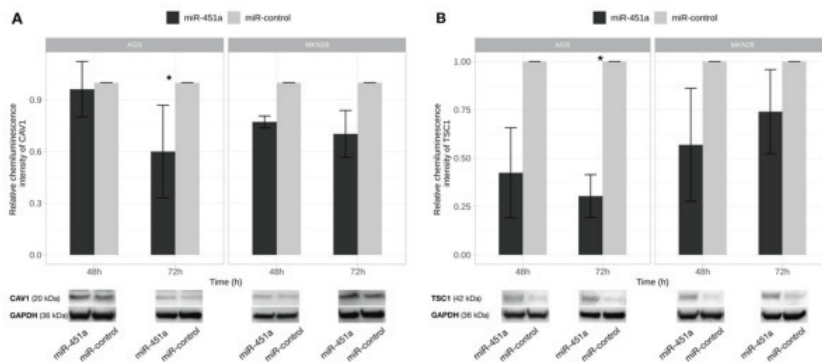
Figure 9. CAV1 and TSC1 proteins level changes after exogenous up-regulation of miR-451a expression. CAV1 (**A**) and TSC1 (**B**) protein expression comparison 48 h and 72 h after transfection in AGS and MKN28 cell cultures transfected with miR-451a and miR-control. Significant CAV1 and TSC1 proteins' level decrease was determined 72 h after transfection in AGS cell culture. Representative pictures of CAV1 and TSC1 proteins signals detected by Western Blot presented at the bottom of a figure (* $p < 0.05$). Data from three independent experiments.

3. Discussion

Although a number of studies have elucidated miRNA profiles of GC, the role of these miRNAs, their possible target genes and functions remain under investigation. In the present study, we analyzed two miRNAs—miR-20b and miR-451a—shown to be deregulated in many different malignancies, including GC [13–16]. We determined that miR-20b and miR-451a directly regulate genes involved in the PI3K/AKT/mTOR signaling pathway, and modification of their expression had a tumor-suppressive role in GC.

MiR-20b belongs to a cluster of highly similar miRNAs called the miR-17 family [22]. Deregulation of miR-20b has been determined in many different cancers [13–15,23–26]. Our results demonstrated that miR-20b is overexpressed in both AGS and MKN28 GC cell lines compared to healthy gastric tissue, which is consistent with previous studies [13,14,27,28]. Moreover, our in vitro results were confirmed in vivo in GC samples and INS-GAS mice, showing increased expression of miR-20b in male mice. The INS-GAS mouse model is gender specific and male gastric tissue responds more rapidly and aggressively to *H. pylori* infection due to differences in hormone secretion. Estrogen may protect female mice from intestinal-type tumors. On the other hand, androgens may promote the development of gastric cancer in male mice [29]. Similar pathogenesis patterns have also been observed in humans [30].

Using loss-of-function experiments we revealed that miR-20b inhibits proliferation and colony formation abilities, reduces viability, and increases ratio of early apoptotic cells in GC cell lines. These results confirm that miR-20b is highly pronounced as oncogenic, driving cellular processes such as cancer cells proliferation [20] and colony formation [31], invasiveness and tumor growth [19]. Our study is the first which investigates the functional importance of miR-20b in gastric carcinogenesis, GC cell lines, and the GC mouse model INS-GAS.

MiR-451a is located in the 17q11.2, region reported to be amplified in GC and other types of cancers [18]. miR-451a is usually down-regulated and acts as a tumor suppressor in a series of cancers [32–34], including GC [16,35]. Our study results have also shown that miR-451a was down-regulated in the investigated GC cell lines. Functional analysis of miR-451a revealed that increased level of this miRNA impaired only colony formation of GC cell lines. A study by Riquelme et al. showed that miR-451a up-regulation in GC could not only affect colony formation but also reduce cell viability [16]. Taking into account other cancer types, increased level of miR-451a has been shown

to impair cell growth, moderately reduce migration ability in thyroid carcinoma cells [32], and inhibit migratory and invasive abilities in renal cell carcinoma [34]. MiR-451a may exert its biological role through mediating tumor formation, maintenance, and metastasis.

We have shown that the inhibition of miR-20b reduced viability, colony formation, and increased ratio of apoptotic cells while increased expression of miR-451a only affected colony formation and was inconsistent with other in vitro assays (MTT and Annexin V/Dead Cell Apoptosis). MTT assay reflects that metabolic activity as well as the colony forming assay is considered to show the ability of cells to proliferate, however both of the assays could also be affected by the broader range of the processes in the cells. Annexin V/Dead Cell Apoptosis assay is specifically based on the Annexin V ability to conjugate with phosphatidylserine, which is usually expressed on the cell surface during apoptosis; however, cells could undergo different molecular changes or different types of cell deaths and result in inconsistency between different in vitro assays.

Also, some results of the loss-of-function assays were not consistent between two commercial gastric cancer cell cultures. This may be caused due to a different origin of cell lines: AGS—primary gastric cancer lesion and MKN28—metastatic site in the liver. This could potentially lead to different characteristics of cells in terms of aggressiveness, invasiveness, proliferative potential, or molecular signatures.

Using in silico analysis, we selected potential target genes of miR-20b: *IRF*, *PTEN,* and *TXNIP* and miR-451a: *CAV1* and *TSC1.* These genes are involved in phosphatidylinositol-3-kinase (PI3K)/AKT/mTOR signaling pathway which has potential prognostic and predictive significance in GC [36,37]. Using Western Blot analysis and luciferase reporter assay we were able to confirm direct *PTEN* and *TXNIP* 3′ UTR interaction with miR-20b and *CAV1* and *TSC1* 3′ UTR interaction with miR-451a.

PTEN is a classical tumor suppressor gene in various human cancers. PTEN functions as a negative regulator of the PI3K/AKT pathway through dephosphorylation of phosphatidylinositol 3,4,5 trisphosphate (PIP3), and is involved in regulation of cellular proliferation, metastasis and apoptosis during progression of cancers [38]. PTEN has been reported to be regulated by numerous miRNAs in multiple cancers, including colorectal carcinoma, glioma, ovarian and breast cancer [31,39–41]. Previously studies investigated *PTEN* as miR-20b target gene in colorectal and breast cancer [26,31], whereas our study is the first to show *PTEN* as an miR-20b target in GC cells.

TXNIP has been identified as potential tumor suppressor gene in various solid tumors and hematological malignancies [42–44]. This gene is involved in PI3K/AKT/mTOR pathway by mediation of glucose intake in cancer cells. The changes in glucose metabolism is associated with a great increase of cell bioenergetic and biosynthetic abilities, which are important to maintain rapid cell proliferation, tumor progression, and resistance to chemotherapy and radiation. Oncogenic activation of PI3K/AKT signaling at least partially promotes cellular glucose uptake through the regulation of TXNIP expression [45]. Previous cancer studies have reported that *TXNIP* is regulated by miR-373 [46], miR-411 [47], and miR-224 [48,49]. Our study is the first to show *TXNIP* as an miR-20b target in GC.

CAV1 (caveolin-1) is a structural component of caveolar membrane domains causing the propagation of downstream signals. Stimulation of CAV1-positive cells results in activation of the PI3K/AKT pathway and leads to cell cycle progression through G1 and entry into the S phase [50]. CAV1 is also involved in regulation of PTEN [51]. Previous studies have revealed several miRNAs involved in direct regulation of *CAV1* [52–54]. However, our study is the first to indicate miR-451a as a direct regulator of *CAV1*.

TSC1 acts through regulation of the mTOR pathway. Activation of upstream signals results in inhibition of TSC1/TSC2 by AKT, allowing mTOR activation. These changes also result in uncontrolled and increased bioenergetic and biosynthetic processes, cell growth, and proliferation [36]. *TSC1* gene as potential target of miR-451a has already been investigated in GC in a study conducted by Riquelme et al. [16]. However, investigators have not confirmed the direct interaction between miR-451a and *TSC1*. Therefore, our study is the first that has confirmed *TSC1* as target gene of miR-451a.

This study has some potential limitations: although commercial cell lines are derived from human primary or metastatic gastric tumor lesion more appropriate control for analyzing miRNA expression patterns would be culture of normal epithelial cells. However, commercial cell cultures showed similar miR-20b and miR-451a expression pattern compared to patients' gastric tumor tissue samples. Further studies analyzing the direct impact of miRNA in vivo using xenographic mouse models would be of interest. Moreover, due to the small sample size we were not able to perform association analysis of miR-20b and miR-451a expression and subphenotypes of GC patients (including *H. pylori* infection).

In conclusion, the present study identified that expressions of miR-20b and miR-451a are significantly deregulated in gastric cancer tissue, commercial cell cultures, and INS-GAS mice. Notably, down-regulation of miR-20b and up-regulation of miR-451a expression exhibits an anti-tumor effect in vitro (reduced viability, colony formation, increased apoptosis rate and reduced colony forming ability, miR-20b, and miR-451a respectively) by targeting genes involved in PI3K/AKT/mTOR tumorigenesis signaling pathway. These findings are important for miRNA functional studies in GC and may also help to improve development of new treatment strategies.

4. Materials and Methods

4.1. Human Tissue Samples and Cell Lines

Study subject recruitment was conducted at the Department of Gastroenterology, Lithuanian University of Health Sciences (Kaunas, Lithuania) where tumor tissue samples ($n = 13$) were obtained from the primary lesion biopsy. The characteristics of patients are listed in Supplementary Table S1. Gastric tissue samples of a control group ($n = 13$) were collected from healthy subjects without atrophy or intestinal metaplasia based on the Sydney classification. All patients in the control group underwent upper endoscopy with biopsies due to dyspeptic symptoms but had no history of malignancy. The samples were stored at $-80\,°C$ as a fresh-frozen sample. The use of biological material in present study was approved by Kaunas Regional Biomedical Research Ethics Committee (protocol no. BE-2-10, 27th May 2011). Each subject has signed written informed consent and all procedures were carried out in accordance with the guidelines of Declaration of Helsinki.

The human gastric adenocarcinoma cell line AGS was obtained from the American Type Culture Collection (ATCC) and MKN28 cell line was kindly provided by Dr. Alexander Link (Department of Gastroenterology, Hepatology and Infectious Diseases, Otto von Guericke University, Germany). Cell cultures were cultivated according to ATCC recommendations. AGS and MKN28 cell lines were cultivated in Ham's F-12K (Kaighn's) Medium (GIBCO Invitrogen Life Technologies, Grand Island, NY, USA) and RPMI 1640 medium (GIBCO Invitrogen Life Technologies, Grand Island, NY, USA), respectively. The culture media was supplemented with 10% Fetal Bovine Serum (FBS) (GIBCO Invitrogen Life Technologies, Grand Island, NY, USA) and 1% penicillin-streptomycin solution (5000 U/mL) (GIBCO Invitrogen Life Technologies, Grand Island, NY, USA). Cells were cultured in humidified incubator containing 5% of CO_2 at 37 °C. Cell lines were tested for mycoplasma contamination using specific primers [55].

4.2. Cell Transfection

MiRNA mimics of miR-451a and miR-20b, miRNA inhibitor of miR-20b (anti-miR-20b), and non-specific miRNA mimic or inhibitor negative control (miRVana[TM], Ambion by Thermo Fisher Scientific, Grand Island, NY, USA) were used for the transfection of AGS and MKN28 cell lines. Transfection was performed using Lipofectamine 3000 transfection reagent (Thermo Fisher Scientific, Waltham, MA, USA) in accordance with the manufacturer's recommendations. A final concentration of 90 nM of anti-miR-20b for AGS cell line, 120 nM of anti-miR-20b for MKN28 cell line and a final concentration of 50 nM of miR-451a mimic for both cell lines was used. Efficiency of transfection was tested and monitored by using positive transfection controls let-7c and miR-1 (miRVana[TM], Ambion by

Thermo Fisher Scientific, Grand Island, NY, USA), and fluorescent siRNA (BLOCK-iT™ Alexa Flour Red Fluorescent, Thermo Fisher Scientific, Waltham, MA, USA).

4.3. Target Prediction

GC-associated putative target-genes of selected miRNAs were retrieved from databases *(DIANA Lab Tools TarBase*, *miRanda*, *TargetScan)* according to their function in carcinogenesis (oncogenes or tumour-suppressor genes). Three potential target genes (*IRF1, TXNIP, PTEN*) with a proto-oncogenic function were selected for miR-20b and two (*CAV1, TSC1*) with an onco-suppressor function for miR-451a.

4.4. qRT-PCR for miR-20b and miR-451a Expression Level

To estimate miRNA expression level total RNA from GC ($n = 13$), normal gastric tissue ($n = 13$), and GC-derived cell lines ($n = 7$) was isolated using miRNeasy Micro Kit (Qiagen, Hilden, Germany) according to manufacturer's protocol. Expression level of miR-20b and miR-451a was determined using TaqMan miRNA assays (Ambion by Thermo Fisher Scientific, Grand Island, NY, USA) using miRNA-specific primers on 7500™ Fast real-time PCR system (Life Technologies, Carlsbad, CA, USA). Levels of miRNA were normalized to miR-16 and changes in expression were calculated using $2^{-\Delta Ct}$ method.

4.5. Western Blot

AGS and MKN28 cells (250,000 and 200,000 cells/well, respectively) were transfected with mimic of miR-451a, inhibitor of miR-20b and respective miRNA controls for 48 h and 72 h. Total protein from cells was lysed in 1× radioimmunoprecipitation assay (RIPA) buffer (Abcam, Cambridge, UK) containing protease and phosphatase inhibitor cocktail (Sigma Aldrich, St. Louis, MO, USA). Protein concentration was estimated using Pierce BCA Protein Assay Kit (Thermo Scientific, USA). Total protein was separated by SDS-PAGE using 4–12% Bis-Tris Plus Mini Gels and blotted to 0.45 μm PVDF membrane. Membranes were blocked in WesternBreeze Blocker/Diluent (Part A and B) (Thermo Fisher Scientific, Waltham, MA, USA) at room temperature for 1 h. Antibodies directed against IRF1 (1:1000 dilution; Cat. No. ab186384; Abcam, Cambridge, UK), PTEN (1:2500 dilution; Cat. No. ab32199, Abcam, Cambridge, UK), TXNIP (8 μg/mL concentration; Cat. No. 40-3700; Thermo Fisher Scientific, Waltham, MA, USA), CAV1 (1:500 dilution, Cat. No. ab192869, Abcam, Cambridge, UK), TSC1 (1:2000 dilution, Cat. No. 37-0400, Thermo Fisher Scientific, Waltham, MA, USA), and GAPDH (0.4 μg/mL concentration; Cat. No. AM4300; Ambion by Thermo Fisher Scientific, Grand Island, NY, USA) were used. The signals were visualized with ChemiDoc XRS+ System (Bio-Rad, Hercules, CA, USA) and ImageLab Software (version 5.2.1, (Bio-Rad, Hercules, CA, USA). Protein levels were normalized to endogenous control of GAPDH protein.

4.6. Luciferase Reporter Assay

Wild-type (wt) and mutant (mut) seed regions of the target genes (*PTEN*, *TXNIP*, *TSC1* and *CAV1*) were constructed and cloned into Luc 3′UTR between the HindIII and BcuI sites of pMIR-REPORT-Luciferase vector (Ambion by Thermo Fisher Scientific, Grand Island, NY, USA). The oligonucleotide sequences are listed in Supplementary Table S2. Constructed vectors were verified by Sanger sequencing using Applied Biosystems® 3500 analyzer (Applied Biosystems, Foster City, CA, USA). AGS cells (100,000 cells/well) were co-transfected with 146 ng of constructs (wt or mut vector) and with 50 nM of either miRNA mimic or negative mimic control using Lipofectamine 3000 (Thermo Fisher Scientific, Waltham, MA, USA). After 48 h incubation luciferase activity was detected by Dual-Light™ Luciferase & β-Galactosidase Reporter Gene Assay System (Applied Biosystems, Foster City, CA, USA) following manufacturer's protocol. Luminescent signal was quantified by GENios Pro microplate reader (Tecan Trading AG, Mannedorf, Switzerland). Reporter activity was normalized to β-Galactosidase activity.

4.7. MTT Assay

Viability and proliferation of AGS and MKN28 cells (4000 and 3500 cells/well, respectively) was measured with 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) assay 48 h and 72 h after transfection. MTT reagent (final concentration 5 mg/mL) (Sigma Aldrich, St. Louis, MO, USA) was added to cells and incubated for 2 h at 37 °C. After incubation supernatant was discarded and formazan was dissolved in 200 µl dimethyl sulfoxide (DMSO) (Sigma Aldrich, St. Louis, MO, USA). Optical density (OD) was determined by Sunrise absorbance microplate reader (Tecan Trading AG, Mannedorf, Switzerland) at wavelength 570 nm and 620 nm (as reference).

4.8. Colony Formation Assay

Both GC cancer cell lines (750 and 1000 cells/well, AGS and MKN28, respectively) were seeded in duplicates onto 6-well plates and transfected after 24 h. After 9 days of incubation of AGS cells and 8 days of incubation of MKN28, cells were washed with PBS, fixed with 10% formalin, and stained with 0.5% crystal violet. Colonies were counted using ImageJ software (version 1.52g).

4.9. Apoptosis Assay

Apoptosis of cells (100,000 cells/well) was measured using FITC Annexin V/Dead Cell Apoptosis Kit with FITC Annexin V and PI (Invitrogen, Carlsbad, CA, USA) 72 h after transfection. Harvested cells were centrifuged and suspended in 1× Annexin binding buffer. Cell suspension was incubated with Annexin V-FITC and PI according to manufacturer's protocol. Samples were analyzed using Accuri C6 flow cytometer (BD Biosciences, Erembodegem, Belgium). Cells were discriminated into viable (both annexin V-FITC/ PI negative), apoptotic (annexin V-FITC positive), and dead cells (both annexin V-FITC/ PI positive).

4.10. Wound Healing Assay

Migration of transfected AGS and MKN28 (100,000 and 75,000 cell/well, respectively) was determined using wound healing assay. Cells were transferred to two well culture-inserts (Ibidi, Munich, Germany) 24 h after transfection. After the overnight incubation, cell culture inserts were removed to create 500 µm gap between the cells. Cell migration was monitored at 0 h, 24 h, and 48 h after the removal of culture-inserts under IX71 microscope (Olympus, Tokyo, Japan).

4.11. Tumorigenicity of miR-20b and miR-451a in INS-GAS Mouse Model

All in vivo experiments were carried out on the insulin-gastrin (INS-GAS) transgenic mouse model. Model was used to evaluate the alterations of miR-20b and miR-451a expression following *H. pylori* infection with a follow-up from 12 to 50 weeks of age.

4.12. Statistical Analysis

Experimental data is presented as means ± standard deviation (SD) of three to five independent experiments. All analyses were performed with R Studio software (R version 3.3.3); a value of $p < 0.05$ was considered statistically significant. Data distribution was determined by the Shapiro–Wilk test, which was used to determine whether the data distribution was normal or non-normal. The *T*-test was used to determine significance of difference between data with normal distribution. The Wilcoxon test was used for significance of difference between data with non-normal distribution.

and S.J.; investigation, G.S., R.I., P.R., S.F. and C.T.; resources, G.K., M.L., J.K., E.K., A.L. and S.S.; writing—original draft preparation, G.S., R.I. and J.S.; writing—review and editing, G.S., J.S. and J.K.; visualization, G.S.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

3′ UTR	three prime untranslated region
ATCC	American type culture collection
CAV1	caveolin-1
DMSO	dimethyl sulfoxide
FBS	fetal bovine serum
FITC	fluorescein isothiocyanate
GAPDH	glyceraldehyde 3-phosphate dehydrogenase
GC	gastric cancer
HER	human epidermal growth factor receptor
INS-GAS	insulin-gastrin
IRF1	interferon regulatory factor 1
miRNA	microRNA
mTOR	mammalian target of rapamycin
MTT	3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide
mut	mutant
OD	optical density
PI	propidium iodide
PI3K	phosphatidylinositol-3-kinase
PTEN	phosphatase and tensin homologue
RIPA	radioimmunoprecipitation assay
RISC	RNA-induced silencing complex
RNA	ribonucleic acid
TSC1	tuberous sclerosis 1
TXNIP	thioredoxin-interacting protein
wt	wild-type

References

1. Ferlay, J.; Soerjomataram, I.; Dikshit, R.; Eser, S.; Mathers, C.; Rebelo, M.; Parkin, D.M.; Forman, D.; Bray, F. Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. *Int. J. Cancer* **2015**, *136*, E359–E386. [CrossRef]
2. Bornschein, J.; Leja, M.; Kupcinskas, J.; Link, A.; Weaver, J.; Rugge, M.; Malfertheiner, P. Molecular diagnostics in gastric cancer. *Front. Biosci. (Landmark Ed)*. **2014**, *19*, 312–338. [CrossRef]
3. Farazi, T.A.; Hoell, J.I.; Morozov, P.; Tuschl, T. microRNAs in Human Cancer. *Adv. Exp. Med. Biol.* **2013**, *774*, 1–20. [PubMed]
4. Bracken, C.P.; Scott, H.S.; Goodall, G.J. A network-biology perspective of microRNA function and dysfunction in cancer. *Nat. Rev. Genet.* **2016**, *17*, 719–732. [CrossRef] [PubMed]
5. Beg, M.S.; Brenner, A.J.; Sachdev, J.; Borad, M.; Kang, Y.-K.; Stoudemire, J.; Smith, S.; Bader, A.G.; Kim, S.; Hong, D.S. Phase I study of MRX34, a liposomal miR-34a mimic, administered twice weekly in patients with advanced solid tumors. *Invest. New Drugs* **2017**, *35*, 180–188. [CrossRef] [PubMed]
6. Demeure, M.J.; Armaghany, T.; Ejadi, S.; Ramanathan, R.K.; Elfiky, A.; Strosberg, J.R.; Smith, D.C.; Whitsett, T.; Liang, W.S.; Sekar, S.; et al. A phase I/II study of TKM-080301, a *PLK1* -targeted RNAi in patients with adrenocortical cancer (ACC). *J. Clin. Oncol.* **2016**, *34*, 2547. [CrossRef]

7. Tolcher, A.W.; Papadopoulos, K.P.; Patnaik, A.; Rasco, D.W.; Martinez, D.; Wood, D.L.; Fielman, B.; Sharma, M.; Janisch, L.A.; Brown, B.D.; et al. Safety and activity of DCR-MYC, a first-in-class Dicer-substrate small interfering RNA (DsiRNA) targeting MYC, in a phase I study in patients with advanced solid tumors. *J. Clin. Oncol.* **2015**, *33*, 11006. [CrossRef]

8. Wagner, M.J.; Mitra, R.; McArthur, M.J.; Baze, W.; Barnhart, K.; Wu, S.Y.; Rodriguez-Aguayo, C.; Zhang, X.; Coleman, R.L.; Lopez-Berestein, G.; et al. Preclinical Mammalian Safety Studies of EPHARNA (DOPC Nanoliposomal EphA2-Targeted siRNA). *Mol. Cancer Ther.* **2017**, *16*, 1114–1123. [CrossRef] [PubMed]

9. Yuan, T.; Huang, X.; Woodcock, M.; Du, M.; Dittmar, R.; Wang, Y.; Tsai, S.; Kohli, M.; Boardman, L.; Patel, T.; et al. Plasma extracellular RNA profiles in healthy and cancer patients. *Sci. Rep.* **2016**, *6*, 19413. [CrossRef]

10. Link, A.; Kupcinskas, J. MicroRNAs as non-invasive diagnostic biomarkers for gastric cancer: Current insights and future perspectives. *World J. Gastroenterol.* **2018**, *24*, 3313–3329. [CrossRef]

11. Juzenas, S.; Salteniene, V.; Kupcinskas, J.; Link, A.; Kiudelis, G.; Jonaitis, L.; Jarmalaite, S.; Kupcinskas, L.; Malfertheiner, P.; Skieceviciene, J. Analysis of deregulated micrornas and their target genes in gastric cancer. *PLoS ONE* **2015**, *10*, e0132327. [CrossRef] [PubMed]

12. Kupcinskas, J.; Wex, T.; Link, A.; Leja, M.; Bruzaite, I.; Steponaitiene, R.; Juzenas, S.; Gyvyte, U.; Ivanauskas, A.; Ancans, G.; et al. Gene polymorphisms of micrornas in helicobacter pylori-induced high risk atrophic gastritis and gastric cancer. *PLoS ONE* **2014**, *9*, e87467. [CrossRef] [PubMed]

13. Xue, T.; Tao, L.; Zhang, M.; Xu, G.; Zhang, J.; Zhang, P. miR-20b overexpression is predictive of poor prognosis in gastric cancer. *Onco. Targets. Ther.* **2015**, *8*, 1871. [CrossRef] [PubMed]

14. Guo, J.; Miao, Y.; Xiao, B.; Huan, R.; Jiang, Z.; Meng, D.; Wang, Y. Differential expression of microRNA species in human gastric cancer versus non-tumorous tissues. *J. Gastroenterol. Hepatol.* **2009**, *24*, 652–657. [CrossRef] [PubMed]

15. Espinosa-Parrilla, Y.; Muñoz, X.; Bonet, C.; Garcia, N.; Venceslá, A.; Yiannakouris, N.; Naccarati, A.; Sieri, S.; Panico, S.; Huerta, J.M.; et al. Genetic association of gastric cancer with miRNA clusters including the cancer-related genes *MIR29, MIR25, MIR93* and *MIR106*: Results from the EPIC-EURGAST study. *Int. J. Cancer* **2014**, *135*, 2065–2076. [CrossRef] [PubMed]

16. Riquelme, I.; Tapia, O.; Leal, P.; Sandoval, A.; Varga, M.G.; Letelier, P.; Buchegger, K.; Bizama, C.; Espinoza, J.A.; Peek, R.M.; et al. miR-101-2, miR-125b-2 and miR-451a act as potential tumor suppressors in gastric cancer through regulation of the PI3K/AKT/mTOR pathway. *Cell. Oncol.* **2016**, *39*, 23–33. [CrossRef] [PubMed]

17. Kiyosawa, N.; Watanabe, K.; Toyama, K.; Ishizuka, H. Circulating miRNA Signature as a Potential Biomarker for the Prediction of Analgesic Efficacy of Hydromorphone. *Int. J. Mol. Sci.* **2019**, *20*, 1665. [CrossRef]

18. Pan, X.; Wang, R.; Wang, Z.-X. The Potential Role of miR-451 in Cancer Diagnosis, Prognosis, and Therapy. *Mol. Cancer Ther.* **2013**, *12*, 1153–1162. [CrossRef]

19. Wang, B.; Yang, J.; Xiao, B. MicroRNA-20b (miR-20b) Promotes the Proliferation, Migration, Invasion, and Tumorigenicity of Esophageal Cancer Cells via the Regulation of Phosphatase and Tensin Homologue Expression. *PLoS ONE* **2016**, *11*, e0164105. [CrossRef]

20. Kawano, M.; Tanaka, K.; Itonaga, I.; Iwasaki, T.; Tsumura, H. MicroRNA-20b promotes cell proliferation via targeting of TGF-β receptor II and upregulates MYC expression in Ewing's sarcoma cells. *Int. J. Oncol.* **2017**, *51*, 1842–1850. [CrossRef]

21. Ulivi, P.; Canale, M.; Passardi, A.; Marisi, G.; Valgiusti, M.; Frassineti, G.L.; Calistri, D.; Amadori, D.; Scarpi, E. Circulating plasma levels of miR-20b, miR-29b and mir-155 as predictors of bevacizumab efficacy in patients with metastatic colorectal cancer. *Int. J. Mol. Sci.* **2018**, *19*, 307. [CrossRef] [PubMed]

22. Tanzer, A.; Stadler, P.F. Molecular Evolution of a MicroRNA Cluster. *J. Mol. Biol.* **2004**, *339*, 327–335. [CrossRef] [PubMed]

23. Guo, J.; Xiao, Z.; Yu, X.; Cao, R. miR-20b promotes cellular proliferation and migration by directly regulating phosphatase and tensin homolog in prostate cancer. *Oncol. Lett.* **2017**, *14*, 6895–6900. [CrossRef] [PubMed]

24. Ahmad, A.; Ginnebaugh, K.R.; Sethi, S.; Chen, W.; Ali, R.; Mittal, S.; Sarkar, F.H. miR-20b is up-regulated in brain metastases from primary breast cancers. *Oncotarget* **2015**, *6*, 12188. [CrossRef] [PubMed]

25. Cheung, T.; Man, K.M.; Yu, M.; Yim, S.; Siu, N.S.S.; Lo, K.W.K.; Doran, G.; Wong, R.R.Y.; Wang, V.W.; Smith, D.I.; et al. Dysregulated microRNAs in the pathogenesis and progression of cervical neoplasm. *Cell Cycle* **2012**, *11*, 2876–2884. [CrossRef]

26.　Zhu, J.; Chen, L.; Zou, L.; Yang, P.; Wu, R.; Mao, Y.; Zhou, H.; Li, R.; Wang, K.; Wang, W.; et al. MiR-20b, -21, and -130b inhibit PTEN expression resulting in B7-H1 over-expression in advanced colorectal cancer. *Hum. Immunol.* **2014**, *75*, 348–353. [CrossRef]

27.　Ishiguro, T.; Ishiguro, H.; Kuwabara, Y.; Kimura, M.; Mitui, A.; Mori, Y.; Ogawa, R.; Harata, K.; Fujii, Y. microRNA expression profile in undifferentiated gastric cancer. *Int. J. Oncol.* **1992**, *34*, 537–542.

28.　Ueda, T.; Volinia, S.; Okumura, H.; Shimizu, M.; Taccioli, C.; Rossi, S.; Alder, H.; Liu, C.; Oue, N.; Yasui, W.; et al. Relation between microRNA expression and progression and prognosis of gastric cancer: A microRNA expression analysis. *Lancet Oncol.* **2010**, *11*, 136–146. [CrossRef]

29.　Fox, J.G.; Wang, T.C.; Rogers, A.B.; Poutahidis, T.; Ge, Z.; Taylor, N.; Dangler, C.A.; Israel, D.A.; Krishna, U.; Gaus, K.; et al. Host and Microbial Constituents Influence Helicobacter pylori-Induced Cancer in a Murine Model of Hypergastrinemia. *Gastroenterology.* **2003**, *124*, 1879–1890. [CrossRef]

30.　Yamamoto, S. Stomach cancer incidence in the world. *Jpn J. Clin. Oncol.* **2001**, *31*, 471–475.

31.　Zhou, W.; Shi, G.; Zhang, Q.; Wu, Q.; Li, B.; Zhang, Z. MicroRNA-20b promotes cell growth of breast cancer cells partly via targeting phosphatase and tensin homologue (PTEN). *Cell Biosci.* **2014**, *4*, 62. [CrossRef] [PubMed]

32.　Minna, E.; Romeo, P.; Dugo, M.; De Cecco, L.; Todoerti, K.; Pilotti, S.; Perrone, F.; Seregni, E.; Agnelli, L.; Neri, A.; et al. miR-451a is underexpressed and targets AKT/mTOR pathway in papillary thyroid carcinoma. *Oncotarget* **2016**, *7*, 12731–12747. [CrossRef]

33.　Sun, H.; Jiang, P. MicroRNA-451a acts as tumor suppressor in cutaneous basal cell carcinoma. *Mol. Genet. Genomic Med.* **2018**, *6*, 1001–1009. [CrossRef] [PubMed]

34.　Yamada, Y.; Arai, T.; Sugawara, S.; Okato, A.; Kato, M.; Kojima, S.; Yamazaki, K.; Naya, Y.; Ichikawa, T.; Seki, N. Impact of novel oncogenic pathways regulated by antitumor *miR-451a* in renal cell carcinoma. *Cancer Sci.* **2018**, *109*, 1239–1253. [CrossRef] [PubMed]

35.　Li, C.-Y.; Liang, G.-Y.; Yao, W.-Z.; Sui, J.; Shen, X.; Zhang, Y.-Q.; Peng, H.; Hong, W.-W.; Ye, Y.-C.; Zhang, Z.-Y.; et al. Identification and functional characterization of microRNAs reveal a potential role in gastric cancer progression. *Clin. Transl. Oncol.* **2017**, *19*, 162–172. [CrossRef] [PubMed]

36.　Tapia, O.; Riquelme, I.; Leal, P.; Sandoval, A.; Aedo, S.; Weber, H.; Letelier, P.; Bellolio, E.; Villaseca, M.; Garcia, P.; et al. The PI3K/AKT/mTOR pathway is activated in gastric cancer with potential prognostic and predictive significance. *Virchows Arch.* **2014**, *465*, 25–33. [CrossRef] [PubMed]

37.　Al-Batran, S.-E.; Ducreux, M.; Ohtsu, A. mTOR as a therapeutic target in patients with gastric cancer. *Int. J. Cancer* **2012**, *130*, 491–496. [CrossRef] [PubMed]

38.　Carnero, A.; Blanco-Aparicio, C.; Renner, O.; Link, W.; Leal, J. The PTEN/PI3K/AKT Signalling Pathway in Cancer, Therapeutic Implications. *Curr. Cancer Drug Targets* **2008**, *8*, 187–198. [CrossRef]

39.　Huse, J.T.; Brennan, C.; Hambardzumyan, D.; Wee, B.; Pena, J.; Rouhanifard, S.H.; Sohn-Lee, C.; le Sage, C.; Agami, R.; Tuschl, T.; et al. The PTEN-regulating microRNA miR-26a is amplified in high-grade glioma and facilitates gliomagenesis in vivo. *Genes Dev.* **2009**, *23*, 1327–1337. [CrossRef]

40.　Wu, W.; Yang, J.; Feng, X.; Wang, H.; Ye, S.; Yang, P.; Tan, W.; Wei, G.; Zhou, Y. MicroRNA-32 (miR-32) regulates phosphatase and tensin homologue (PTEN) expression and promotes growth, migration, and invasion in colorectal carcinoma cells. *Mol. Cancer* **2013**, *12*, 30. [CrossRef]

41.　Yang, H.; Kong, W.; He, L.; Zhao, J.-J.; O'Donnell, J.D.; Wang, J.; Wenham, R.M.; Coppola, D.; Kruk, P.A.; Nicosia, S.V.; et al. MicroRNA Expression Profiling in Human Ovarian Cancer: miR-214 Induces Cell Survival and Cisplatin Resistance by Targeting PTEN. *Cancer Res.* **2008**, *68*, 425–433. [CrossRef] [PubMed]

42.　Zhou, J.; Yu, Q.; Chng, W.-J. TXNIP (VDUP-1, TBP-2): A major redox regulator commonly suppressed in cancer by epigenetic mechanisms. *Int. J. Biochem. Cell Biol.* **2011**, *43*, 1668–1673. [CrossRef] [PubMed]

43.　Zhou, J.; Chng, W.-J. Roles of thioredoxin binding protein (TXNIP) in oxidative stress, apoptosis and cancer. *Mitochondrion* **2013**, *13*, 163–169. [CrossRef]

44.　Yoshihara, E.; Masaki, S.; Matsuo, Y.; Chen, Z.; Tian, H.; Yodoi, J. Thioredoxin/Txnip: Redoxisome, as a Redox Switch for the Pathogenesis of Diseases. *Front. Immunol.* **2014**, *4*, 514. [CrossRef] [PubMed]

45.　Hong, S.Y.; Yu, F.-X.; Luo, Y.; Hagen, T. Oncogenic activation of the PI3K/Akt pathway promotes cellular glucose uptake by downregulating the expression of thioredoxin-interacting protein. *Cell. Signal.* **2016**, *28*, 377–383. [CrossRef] [PubMed]

46. Chen, D.; Dang, B.-L.; Huang, J.; Chen, M.; Wu, D.; Xu, M.-L.; Li, R.; Yan, G.R. MiR-373 drives the epithelial-to-mesenchymal transition and metastasis via the miR-373-TXNIP-HIF1α-TWIST signaling axis in breast cancer. *Oncotarget* **2015**, *6*, 32701–32712.

47. Zhang, C.; Wang, H.; Liu, X.; Hu, Y.; Ding, L.; Zhang, X.; Sun, Q.; Li, Y. Oncogenic microRNA-411 promotes lung carcinogenesis by directly targeting suppressor genes SPRY4 and TXNIP. *Oncogene* **2019**, *38*, 1892–1904. [CrossRef]

48. Zhu, G.; Zhou, L.; Liu, H.; Shan, Y.; Zhang, X. MicroRNA-224 Promotes Pancreatic Cancer Cell Proliferation and Migration by Targeting the TXNIP-Mediated HIF1α Pathway. *Cell. Physiol. Biochem.* **2018**, *48*, 1735–1746. [CrossRef]

49. Knoll, S.; Fürst, K.; Kowtharapu, B.; Schmitz, U.; Marquardt, S.; Wolkenhauer, O.; Martin, H.; Pützer, B.M. E2F1 induces miR-224/452 expression to drive EMT through TXNIP downregulation. *EMBO Rep.* **2014**, *15*, 1315–1329. [CrossRef]

50. Park, J.H.; Han, H.J. Caveolin-1 plays important role in EGF-induced migration and proliferation of mouse embryonic stem cells: involvement of PI3K/Akt and ERK. *Am. J. Physiol. Physiol.* **2009**, *297*, C935–C944. [CrossRef]

51. Caselli, A.; Mazzinghi, B.; Camici, G.; Manao, G.; Ramponi, G. Some protein tyrosine phosphatases target in part to lipid rafts and interact with caveolin-1. *Biochem. Biophys. Res. Commun.* **2002**, *296*, 692–697. [CrossRef]

52. Zhou, W.; He, L.; Dai, Y.; Zhang, Y.; Wang, J.; Liu, B. MicroRNA-124 inhibits cell proliferation, invasion and migration by targeting CAV1 in bladder cancer. *Exp. Ther. Med.* **2018**, *16*, 2811. [CrossRef] [PubMed]

53. Liu, B.; Zhang, J.; Yang, D. miR-96-5p promotes the proliferation and migration of ovarian cancer cells by suppressing Caveolae1. *J. Ovarian Res.* **2019**, *12*, 57. [CrossRef] [PubMed]

54. Kanlikilicer, P.; Bayraktar, R.; Denizli, M.; Rashed, M.H.; Ivan, C.; Aslan, B.; Mitra, R.; Karagoz, K.; Bayraktar, E.; Zhang, X.; et al. Exosomal miRNA confers chemo resistance via targeting Cav1/p-gp/M2-type macrophage axis in ovarian cancer. *EBioMedicine* **2018**, *38*, 100–112. [CrossRef]

55. Young, L.; Sung, J.; Stacey, G.; Masters, J.R. Detection of Mycoplasma in cell cultures. *Nat. Protoc.* **2010**, *5*, 929–934. [CrossRef]

A3

Title: Quantifying cell free DNA in urine: comparison between commercial kits, impact of gender and inter-individual variation

Authors: Streleckienė, Greta; Reid, Hayley M; Arnold, Norbert; Bauerschlag, Dirk; Forster, Michael

BioTechniques 64, 5 (2018)

Quantifying cell-free DNA in urine: comparison between commercial kits, impact of gender and inter-individual variation

Greta Streleckiene[1,2], Hayley M Reid[2], Norbert Arnold[2,3], Dirk Bauerschlag[3] & Michael Forster*[,2]

[1]Institute for Digestive Research, Lithuanian University of Health Sciences, Eiveniu Str. 2, LT-50009 Kaunas, Lithuania, [2]Institute of Clinical Molecular Biology, Christian-Albrechts-Universität zu Kiel, Schleswig-Holstein, D-24105 Kiel, Germany, [3]Department of Gynecology & Obstetrics, University Hospital of Schleswig-Holstein, Christian-Albrechts-Universität zu Kiel, D-24105 Kiel, Germany

DNA can enter the blood circulation from living cells by extracellular vesicles or at cell death, and pass into urine through the kidney barrier. Urine can be collected non-invasively, making it an interesting source of cell-free DNA (cfDNA) for research studies and ultimately for clinical diagnostics. However, there is currently a lack of data on the quantity and variability of cfDNA in urine. Here, we benchmark two commercial urine cfDNA isolation kits with respect to the quantity of DNA, the labor time, and cost. The results show distinctive differences between each kit. Furthermore, the cfDNA amount from the same probands varied strongly from day to day and may be higher in female samples than in male samples (p = 0.003).

Cell-free DNA is released from cells at cell death and also through active cell secretion [1–3]. cfDNA is readily available in various body fluids, such as blood plasma and urine [4–6]. One advantage of urine over blood is non-invasive sampling at the study probands' or patients' own choice of time, without the need for medical staff. Urine sampling could improve proband recruitment success and compliance, and avoids potential ethical problems in studies that involve children or patients in great pain. In contrast to blood cfDNA, there are insufficient data on the quantity and variability of cfDNA in urine. This fundamental information is required for designing urine cfDNA studies.

Here, we benchmarked two new commercial urine cfDNA isolation kits from Norgen and PerkinElmer with respect to cfDNA yields, replicability of yields from the same probands' urine at sequential intervals, and potential differences between males and females. We chose these specific urine kits because the Norgen blood plasma cfDNA isolation kits showed the highest yields in a benchmark [7] and the Perkin-Elmer plasma cfDNA kit yields in our lab were even higher (data not shown).

Informed consent was obtained from all participants, consisting of healthy volunteers and cancer patients. The University Hospital of Schleswig-Holstein's ethics committee approved the patient information sheet and the consent form used for the study (B327/10, D470/14). All patients included in the study gave written informed consent to donate their samples to the biobank for research use. The research was conducted according to the principles of the Declaration of Helsinki. Initially, we collected samples from cancer patients and healthy volunteers in Urine Collection and Preservation Tubes (15 cc) (Norgen Biotek, Cat.

18120) and isolated cfDNA using the Perkin-Elmer kit and the Norgen kit. The cfDNA yields ranged from 0 to 86.5 ng cfDNA from 5 ml of urine (data not shown). It was unclear whether this variability was of biological nature.

Therefore, we systematically compared the kits by examining the reproducibility of urine cfDNA yield from eight healthy volunteers (four females and four males) on 5 different days. They were informed that the second morning urine was required, as the first morning urine was reported to contain more degraded cfDNA [8]. 100 ml sterile disposible urine cups were distributed. Urine donation took place between 9am and 10am on each day of donation. No preservation agents were used. Urine processing started between 10am and 11am on the same day. The urine samples were centrifuged twice to remove cellular matter, first at 200 x g (10 min) followed by 1800

METHOD SUMMARY

Two new kits were benchmarked for urine cfDNA isolation: magnetic bead based isolation (NEXTprep-Mag Urine cfDNA Isolation Kit, Cat. NOVA-3826–02, PerkinElmer, MA, USA) and silica gel membrane column-based isolation (Urine Cell-Free Circulating DNA Purification Midi Kit, Cat. 56700, Norgen Biotek, ON, Canada). The quantity and length distribution of cfDNA were evaluated by an automated electrophoresis system (Agilent 2200 TapeStation using the High Sensitivity D5000 Screen Tape, Cat. 5067–5588, Agilent Technologies, Waldbronn, Germany).
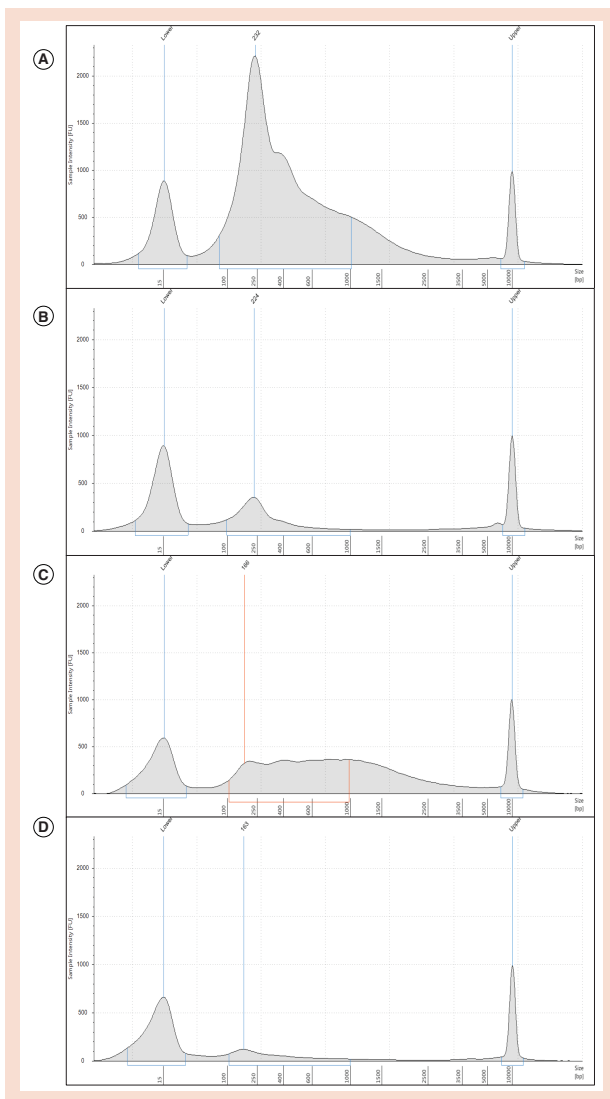
107

Figure 1. Urine cfDNA length distributions depending on gender and kit. The figure shows Agilent TapeStation electropherograms of cfDNA from pooled urine samples from the fourth donation day. **(A)** Isolated with PerkinElmer kit from pooled healthy female urine. **(B)** Isolated with PerkinElmer kit from pooled healthy male urine. **(C)** Isolated with Norgen kit from same pooled female urine. **(D)** Isolated with Norgen Kit from same pooled male urine. The DNA fragment size range from 100–1000 base pairs was manually selected for the quantification of the cfDNA yields.

x g (10 min). Equal volumes from the four female or four male samples were pooled

to obtain an average result for female or male donors without the cost of isolating

each individual's cfDNA. 10 ml of each pool was used for Norgen-based cfDNA isolation and 4 ml for PerkinElmer-based isolation. Additionally, on the fifth day of donation, cfDNA was isolated from each individual's urine sample (F1–F4 and M1–M4). All isolations were performed according to manufacturers' instructions. Length distribution and concentration of isolated DNA were analyzed using the Agilent TapeStation, according to the manufacturer's instructions. A DNA size range of 100 to 1000 bp was manually selected in the TapeStation Analysis Software A.01.04 to obtain the cfDNA concentration.

Figure 1 compares electropherograms between isolates from the same pooled female and male urine samples, respectively. Figure 2A and Table 1 summarize the cfDNA yield ranges and differences between kits in the female and male pools on five donation days. The mean yields of cfDNA (ng per ml of urine ± standard deviation) were 4.72 ± 3.53 (Norgen) and 14.83 ± 11.16 (PerkinElmer) for female pools and 0.83 ± 0.39 (Norgen) and 2.04 ± 0.33 (PerkinElmer) for male pools. The yield difference between kits was significant for male pools ($p = 0.008$, Wilcoxon test). The yield difference between genders was significant for the Norgen kit ($p = 0.016$, Wilcoxon test) and when both kits were considered ($p = 0.003$, Wilcoxon test). Figure 2B and Table 2 indicate yield ranges and kit-based differences for the eight probands' individual urine samples that were isolated on the fifth day. The mean yields of cfDNA (ng per ml of urine ± standard deviation) were 0.98 ± 0.40 (Norgen) and 2.27 ± 1.17 (PerkinElmer) for female individuals and 1.46 ± 0.67 (Norgen) and 1.99 ± 1.02 (PerkinElmer) for male individuals. Finally, Table 3 summarizes the isolation cost and processing times.

In conclusion, the bead-based method was twice as fast as the column-based method and tended to yield more cfDNA per ml of urine. Larger sample numbers may lead to greater clarity. The urine cfDNA length profiles (Figure 1) suggest that the PerkinElmer kit is more efficient at capturing short DNA. Short DNA is of scientific interest, as it is present in blood plasma as cfDNA. In blood plasma, cfDNA lengths peak at about 165 nucleotides with a minor peak at about 1000 nucleotides [9]. In urine, we observed prominent DNA fractions longer than 165 nucleotides. These fractions of longer DNA fragments possibly arise from the epithelia of the urinary tract, after shedding
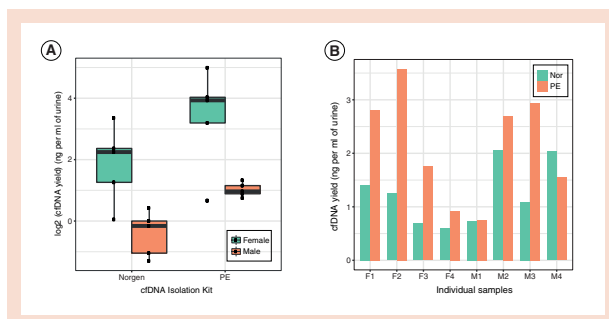
Figure 2. Urine cfDNA isolation yields depending on kit and individual. (A) Pooled urine isolations. Norgen and PerkinElmer cfDNA isolation yields from the same healthy female and male urine pools replicated on five days. Yield varies strongly within the same groups of study probands between different days. Yield from the same urine is generally higher with the PerkinElmer kit. Yield is higher in the female pools. **(B)** Individuals' urine isolations. Norgen and PerkinElmer cfDNA isolation yields from four healthy female probands' individual urine samples (F1-F4) and from four healthy male probands' individual urine samples (M1–M4). The PerkinElmer kit yields are higher than the Norgen kit yields, especially for the female urine. Fourfold inter-individual differences are seen between lowest and highest yield.

(7 extractions × 10 ml using the Midi kit or 3 × 30 ml using the Maxi kit).

Author contributions

GS, HR and MF prepared the manuscript. GS and HR performed cfDNA isolations, cfDNA library preparation and sequencing experiments. GS and MF analyzed the data. NA and DB provided cancer patient samples and helped write the manuscript. All authors read and approved the final version.

Acknowledgements

and subsequent lysis of epithelial cells. Our gender-related differences in cfDNA yields correspond to a previous study on genomic DNA in urine [10,11]. Higher DNA yields in female urine have been previously reported [11–13] indicating that urine from females contained more epithelial cells than that from males.

To sum up, there is vast variation in DNA yields between different individuals and even for the same individuals on different days. When designing a study, we recommend ample amounts of urine to be collected. For NGS with 50 ng of cfDNA, we recommend collecting 60 ml urine for the PerkinElmer kit (15 extractions × 4 ml) or 70 ml for Norgen

Ethical conduct of research

Informed consent was obtained from all participants, consisting of healthy volunteers and cancer patients. The University Hospital of Schleswig-Holstein's ethics committee approved the patient information sheet and the consent form used for the study (B327/10, D470/14). All patients included in the study gave written informed consent

Table 1. cfDNA yields from female and male urine sample pools† on five different collection days.

Donation day	Isolation kit	Gender of pool	cfDNA (ng per ml of urine)	cfDNA per urine isolation (ng)‡
1.	Norgen	F	4.75	47.45
		M	1.00	10.00
	PerkinElmer	F	15.25	61.00
		M	1.68	6.72
2.	Norgen	F	10.25	102.50
		M	0.48	4.85
	PerkinElmer	F	31.80	127.20
		M	2.23	8.90
3.	Norgen	F	5.15	51.50
		M	0.90	8.95
	PerkinElmer	F	16.35	65.40
		M	1.85	7.40
4.	Norgen	F	2.40	23.95
		M	0.41	4.07
	PerkinElmer	F	9.15	36.60
		M	2.51	10.04
5.	Norgen	F	1.04	10.37
		M	1.34	13.42
	PerkinElmer	F	1.58	6.34
		M	1.95	7.81

†Female urine pools from the same four healthy females, male urine pools from the same four healthy males.
‡Urine volumes in the cfDNA isolation protocols are 10 ml (Norgen) vs 4 ml (Perkin Elmer).

Table 2. cfDNA isolation results from eight healthy individuals' urine samples.

Gender	Sample	Isolation kit	cfDNA (ng per ml of urine)	cfDNA per urine sample (ng)[†]
Female	F1	Norgen	1.40	14.03
		PerkinElmer	2.81	11.22
	F2	Norgen	1.25	12.49
		PerkinElmer	3.58	14.30
	F3	Norgen	0.69	6.90
		PerkinElmer	1.75	7.01
	F4	Norgen	0.59	5.94
		PerkinElmer	0.92	3.70
Male	**M1**	**Norgen**	**0.73**	**7.32**
		PerkinElmer	**0.76**	**3.03**
	M2	**Norgen**	**2.05**	**20.54**
		PerkinElmer	**2.70**	**10.79**
	M3	**Norgen**	**1.09**	**10.89**
		PerkinElmer	**2.94**	**11.77**
	M4	**Norgen**	**2.04**	**20.35**
		PerkinElmer	**1.55**	**6.20**

[†]Urine volumes in the cfDNA isolation protocols are 10 ml (Norgen) vs 4 ml (Perkin Elmer).

Table 3. Commercial cfDNA isolation kits used in this study.

Full name of kit	Manufacturer	Urine amount (ml)	Elution volume (μl)	Price per sample (€)	Processing time (min)
Urine Cell-Free Circulating DNA Purification Midi Kit	Norgen Biotek	10	50	20.50	90
NextPrep-Mag Urine cfDNA Isolation Kit	Bioo Scientific by Perkin Elmer	4	20	12.17	45

to donate their samples to the biobank for research use. The research was conducted according to the principles of the Declaration of Helsinki.

Financial & competing interests disclosure

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties. No writing assistance was utilized in the production of this manuscript.

Open access

This work is licensed under the Creative Commons Attribution 4.0 License. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/

References

1. Thierry AR, El Messaoudi S, Gahan PB, Anker P, Stroun M. Origins, structures, and functions of circulating DNA in oncology. *Cancer Metastasis Rev.* 35(3), 347–376 (2016).
2. Thakur BK, Zhang H, Becke A Double-stranded DNA in exosomes: a novel biomarker in cancer detection. *Cell Res.* 24(6), 766–769 (2014).
3. Minciacchi VR, Zijlstra A, Rubin MA, Di Vizio D. Extracellular vesicles for liquid biopsy in prostate cancer: where are we and where are we headed? *Prostate Cancer Prostatic Dis.* 20(3), 251–258 (2017).
4. Botezatu I, Serdyuk O, Potapova G Genetic analysis of DNA excreted in urine: a new approach for detecting specific genomic DNA sequences from cells dying in an organism. *Clin. Chem.* 46(8 Pt 1), 1078–1084 (2000).
5. Chan KCA, Leung SF, Yeung SW, Chan ATC, Lo YMD. Quantitative analysis of the transrenal excretion of circulating EBV DNA in nasopharyngeal carcinoma patients. *Clin. Cancer Res.* 14(15), 4809–4813 (2008).
6. Birkenkamp-Demtröder K, Nordentoft I, Christensen E Genomic alterations in liquid biopsies from patients with bladder cancer. *Eur. Urol.* 70(1), 75–82 (2016).
7. Mauger F, Dulary C, Daviaud C, Deleuze JF, Tost J. Comprehensive evaluation of methods to isolate, quantify, and characterize circulating cell-free DNA from small volumes of plasma. *Anal. Bioanal. Chem.* 407, 6873–6878 (2015).
8. Brisuda A, Pazourkova E, Soukup V Urinary cell-free DNA quantification as non-invasive biomarker in patients with bladder cancer. *Urol. Int.* 96(1), 25–31 (2016).
9. Jiang P, Lo YMD. The long and short of circulating cell-free DNA and the ins and outs of molecular diagnostics. *Trends Genet.* 32(6), 360–371 (2016).
10. El Bali L, Diman A, Bernard A, Roosens NHC, De Keersmaecker SCJ. Comparative study of seven commercial kits for human dna extraction from urine samples suitable for DNA biomarker-based public health studies. *J. Biomol. Tech.* 25(4), 96–110 (2014).
11. Chiarella P, Carbonari D, Capone P Susceptibility biomarker detection in urine exfoliate DNA. *Biomark. Med.* 11(11), 957–966 (2017).
12. Vu NT, Chaturvedi AK, Canfield DV. Genotyping for DQA1 and PM loci in urine using PCR-based amplification: effects of sample volume, storage temperature, preservatives, and aging on DNA extraction and typing. *Forensic Sci. Int.* 102(1), 23–34 (1999).
13. Nakazono T, Kashimura S, Hayashiba Y, Hara K, Miyoshi A. Successful DNA typing of urine stains using a DNA purification kit following dialfiltration. *J. Forensic Sci.* 50(4), 860–864 (2005).

To purchase reprints of this article, contact s.cavana@future-science.com

Title: Effects of quantification methods, isolation kits, plasma biobanking, and hemolysis on cell-free DNA analysis in plasma

Authors: Streleckienė, Greta; Forster, Michael; Inčiūraitė, Rūta; Lukoševičius, Rokas; Skiecevičienė, Jurgita

Biopreservation and biobanking 17, 6 (2019)

Reprinted under licence agreement no. 1191309 (Supplement 5).

Effects of Quantification Methods, Isolation Kits, Plasma Biobanking, and Hemolysis on Cell-Free DNA Analysis in Plasma

Greta Streleckiene,[1] Michael Forster,[2] Ruta Inciuraite,[1] Rokas Lukosevicius,[1] and Jurgita Skieceviciene[1]

Cell-free DNA (cfDNA) has become a promising noninvasive clinical marker widely studied in early disease detection, monitoring, and therapy selection. However, there is lack of data on a number of cfDNA-associated procedural features such as blood plasma biobanking conditions, isolation, and quantification methods that should be taken into account as they can affect downstream applications. In this study cfDNA from 125 plasma samples from healthy individuals were isolated using three different commercial kits (bead and vacuum based). Yield of cfDNA, distribution of cfDNA fragments and absolute amount of miR-223 were estimated. Moreover, the impact of different plasma biobanking conditions and hemolytic plasma were evaluated. In conclusion, results showed that quantification method (fluorescence or microcapillary electrophoresis based) has a major impact in estimating cfDNA amount. Samples isolated by QIAamp showed a higher amount of larger (around 300 bp) DNA fragments and miRNA yield, suggesting possible applications for multiomics approach. On the other hand, the highest cfDNA yield was obtained in samples isolated by the MagMAX Isolation Kit. This kit also showed lowest coefficient of variation and low miRNA yield. Plasma storage conditions and hemolysis affected performance of isolation kits differently.

Keywords: cell-free DNA, plasma, quantification, plasma biobanking, hemolysis

Introduction

CELL-FREE DNA (cfDNA) is released into the blood stream in various ways, including at the death of cells or by active secretion.[1] cfDNA molecules are small with a major peak at 160–180 bp.[2] Increased levels of cfDNA can be detected in various physiological and pathological conditions.[3] Analysis of cfDNA in the blood of cancer patients showed a strong correlation between cfDNA amount and tumor size, burden, and metastasis status. It was also shown that cfDNA could be a powerful disease state and relapse monitoring analyte.[1,4–6]

New minimally invasive diagnostic procedures for circulating molecules are in demand because standard diagnostics are not able to analyze the cancer mutation profile changes over the course of treatment. Apart from cancer mutation monitoring, circulating nucleic acids could be used for multiomic analysis comprising genomic and epigenetic (miRNA) alterations for various pathological conditions. Therefore, blood plasma is a compelling source for liquid biopsies and there is a high interest among the medical and biobanking communities for liquid biopsy diagnostics and disease monitoring. However, there is currently a lack of scientific reports on standardized procedures for plasma storage conditions, handling of hemolyzed plasma, quantification and cfDNA isolation methods, and a lack of reports on analyzing both circulating DNA and miRNA yields.

Previous studies analyzed the impact of stabilizing agents (specifically for cfDNA and routinely used stabilizers [EDTA, heparin, etc.]),[7–11] the time elapsed between blood collection and centrifugation,[9,11,12] different centrifugation profiles, and isolation kits.[8,11,13,14] In the study we present here, we evaluated differences between two quantification methods and three cfDNA isolation kits. Additionally, we analyzed whether plasma storage conditions and hemolysis affect the amount of isolated cfDNA. The key metrics in our analyses were the yields of genomic DNA (gDNA) and miRNA in cfDNA samples and the size distribution of smaller and larger cfDNA fragments between different kits. The isolation kits were selected to compare the performance of different methods (bead and vacuum based).

[1]Institute for Digestive Research, Lithuanian University of Health Sciences, Kaunas, Lithuania.
[2]Institute of Clinical Molecular Biology, Christian-Albrechts University of Kiel, Kiel, Germany.

1

To our knowledge this is the first study in which the discrepancy between cfDNA quantification methods, the impact of hemolysis on cfDNA amount, and the amount of miRNA in cfDNA samples (multiomics approach) were evaluated.

Materials and Methods

The study was approved by the Kaunas Regional Biomedical Research Ethics Committee (No. Nr. BE-2-10). Informed consent was obtained from all study participants.

Blood collection, plasma storage, and hemolysis induction

Blood was collected by standard phlebotomy techniques in BD Vacutainer K_2EDTA tubes (10 mL, 366643) at two time points during the period of the study. For studying the influence of plasma freeze–thaw cycles, our experiment was repeated on two separate days for the same volunteers (age range of subjects were 22–25 years) and set up as follows (see Fig. 1, left half): three blood tubes were collected from five self-reported healthy volunteers (one male and four females). Blood was processed immediately after blood draw to avoid confounding effects.

Tubes were centrifuged at room temperature at $2000 \times g$ for 10 minutes with smooth brake profile to prevent disruption of the buffy coat layer. Equal volumes (6 mL) of each participant's plasma were pooled and divided into three groups: (i) fresh plasma, (ii) one freeze–thaw cycle, and (iii) two freeze–thaw cycles.

The fresh plasma aliquots from group (i) were centrifuged again at $3000 \times g$ for 10 minutes with smooth brake profile according to the recommendations from previous studies.[7,9] The aliquots from group (ii) (one freeze–thaw cycle) and group (iii) (two freeze–thaw cycles) were frozen at $-20°C$ after just one blood centrifugation. The aliquots from group (iii) (two freeze–thaw cycles) were then completely thawed at room temperature and frozen at $-20°C$ once again. All frozen plasma samples were centrifuged before cfDNA isolation procedures at 4°C at $16,000 \times g$ for 10 minutes according to the isolation kit manufacturer's recommendations.

For studying the influence of hemolysis, our experiment was repeated on two separate days for the same volunteer and set up as follows (see Fig. 1, right half): three tubes of blood were collected from one healthy volunteer. Of these tubes, one nonhemolyzed blood tube was centrifuged at room temperature at $2000 \times g$ for 10 minutes with smooth brake profile and then at $3000 \times g$ for 10 minutes with smooth brake profile.

Hemolysis was induced mechanically on the remaining two blood tubes as described by Koseoglu et al.[15]: one and five draws through the needle were performed to obtain moderate and severe hemolysis, respectively. Then blood samples were first centrifuged with smooth brake profile at $2000 \times g$ for 10 minutes and second at $3000 \times g$ for 10 minutes. Finally, plasma samples were aliquoted (2×1 mL) of plasma for each tested condition: nonhemolyzed, $1 \times$ hemolyzed, and $5 \times$ hemolyzed) and processed further. Hemolysis of blood samples was evaluated spectrophotometrically at 570 nm wavelength using an absorbance microplate reader, Tecan Sunrise[16] (data not shown).
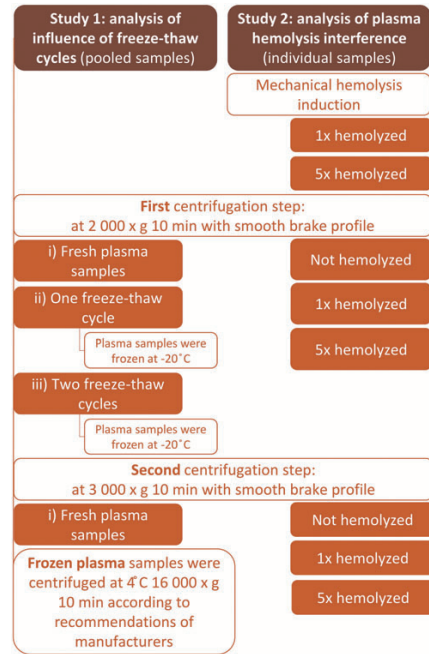

FIG. 1. Flow of samples preparation. Study 1: pooled samples from five healthy volunteers; Study 2: individual samples from one healthy volunteer. *Left half:* equal volumes of each participant's plasma were pooled and divided into three groups: (i) fresh plasma, (ii) one freeze–thaw cycle, and (iii) two freeze–thaw cycles. The blood tubes were first centrifuged at room temperature at $2000 \times g$ for 10 minutes. Then fresh plasma aliquots from group (i) were centrifuged again at $3000 \times g$ for 10 minutes. The aliquots from group (ii) (one freeze–thaw cycle) and group (iii) (two freeze–thaw cycles) were frozen at $-20°C$ after just one blood centrifugation. The aliquots from group (iii) (two freeze–thaw cycles) were then completely thawed at room temperature and frozen at $-20°C$ once again. All frozen plasma samples were centrifuged before cfDNA isolation procedures at 4°C at $16,000 \times g$ for 10 minutes. The plasma storage experiment was repeated in another two replicates. *Right half:* Three blood tubes were drawn from a healthy proband. The first tube of nonhemolyzed blood was centrifuged at room temperature at $2000 \times g$ for 10 minutes and then at $3000 \times g$ for 10 minutes. Hemolysis was induced mechanically on the remaining two blood tubes: one and five draws through the needle were performed. The hemolysis experiment was repeated in another three replicates. cfDNA, cell-free DNA.

cfDNA isolation

cfDNA was isolated using three commercially available isolation kits. The NextPrep-Mag cfDNA Isolation Kit (Bioo Scientific) and MagMAX Cell-Free DNA Isolation

TABLE 1. SPECIFICATIONS OF CELL-FREE DNA ISOLATION KITS

Isolation kit	Manufacturer	Plasma input	Price per 1 mL of plasma (Eur)[a]	Price per 5 mL of plasma (Eur)[a]	Technique	Carrier RNA	Proteinase K
NextPrep-Mag	Bioo Scientific by PerkinElmer	<1–3 mL	9.73	48.67[b]	Beads-based	No	Yes
MagMAX	Thermo Scientific	500 μL–10 mL	11.47	57.34	Beads-based	No	Optional
QIAamp	Qiagen	1–5 mL	22.33	22.33	Vacuum-based	Yes	Yes

[a]Price may vary in different countries.
[b]Price for the here tested 1–3 mL kit; the 3–5 mL kit is better value for 5 mL isolations.

Kit (Thermo Fisher Scientific) are bead based, whereas the QIAamp Circulating Nucleic Acid Kit (Qiagen) is a column-based isolation kit (Table 1). All isolation procedures were done according to the manufacturer's protocols using 1 mL of plasma. To assess exogenous DNA contamination, a water control was included. Purified DNA was either analyzed without delay or stored at +4 until further analysis. Processing time of the cfDNA Isolation Kits was around 45 minutes, 1.5 hours, and 2 hours for NextPrep-Mag, MagMAX, and QIAamp, respectively.

cfDNA quantification

cfDNA amount was quantified by two different methods: the Fluorescence-based Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific) and laser-induced fluorescence-based microcapillary electrophoresis, the Agilent High-Sensitivity DNA Kit (Agilent Technologies). Peaks around 180, 350, and 500 bp were considered as mono-, di-, and trinucleosomal fragments, respectively (Fig. 2). Fragments higher than 1000 bases were considered as gDNA. Summary statistics are presented in Table 2.

Absolute miR-223 quantification

The amount of miR-223 in plasma samples was determined using real-time quantitative polymerase chain reaction (RT-qPCR) and TaqMan miRNA Assay primers and probes. RT-qPCR was carried out on Applied Biosystems 7500 Fast Thermocycler using the manufacturer's recommended cycling conditions. Absolute quantification was performed using qPCR standard curves generated with synthetic single-stranded RNA nucleotides corresponding to the mature miR-223 sequence (miRbase Release v.22). The synthetic miRNA was first reverse transcribed to generate standard curves, then the miRNA amount in pg per 1 mL of plasma was evaluated according to the C_T value in each sample.

Statistical analysis

Statistical analysis was performed using R Studio software (R version 3.3.3). Data are presented as mean ± standard deviation and considered significant when p is less than 0.05. Two-sided T-test or Mann–Whitney U test were used. Data distribution was determined by the Kolmogorov–Smirnov (K–S) test, which was used to determine whether the data distribution was normal or non-normal. The T-test was used to determine significance of difference between data with normal distribution. The Mann–Whitney U test was used for significance of difference between data with non-normal distribution.

Results

Amount of cfDNA varies greatly depending on quantification method

First, our study determined whether there are differences between values obtained by fluorescence-based and laser-induced fluorescence-based microcapillary electrophoresis methods (Qubit and Agilent, respectively) for all samples of the study ($n = 125$). Interestingly, values obtained by the two different methods differed significantly (on average Qubit: 7.11 ± 3.24 ng/mL of plasma; Agilent: 4.06 ± 1.55 ng/mL of plasma; Mann–Whitney U test, $p = 1.86 \times 10^{-21}$). Accordingly, comparing different kits, differences in the amount of cfDNA were significant only on the basis of the Qubit quantification method (Fig. 3): the amount of cfDNA isolated by QIAamp was significantly greater (9.49 ± 4.02 ng/mL of plasma) compared with the NextPrep-Mag (5.74 ± 1.57 ng/mL of plasma) and MagMAX (6.06 ± 2.08 ng/mL of plasma) Isolation Kits (Mann–Whitney U test, $p = 1.62 \times 10^{-7}$; $p = 1.16 \times 10^{-6}$, respectively).

Conversely, according to the Agilent quantification method, there were no significant differences between the amounts of cfDNA isolated with different isolation kits (Fig. 3). The discrepancy between quantification methods may be explained by the lack of specificity for cfDNA in fluorescence-based methods, as not only cfDNA fragments are quantified. Carrier RNA (cRNA) is present in QIAamp isolations, which could interfere with the signal. Therefore, all further comparisons were based on quantifications by the Agilent method. The coefficient of variation (CV) of cfDNA yield in aliquoted pooled samples quantified by Qubit and Agilent isolated by NextPrep-Mag were—12.32% and 30.05%, MagMAX—11.68% and 14.12%, and QIAamp—14.91% and 26.30%, respectively (Table 2).

Plasma biobanking conditions and hemolysis affected performance of cfDNA isolation kits differently

We have evaluated how freeze–thaw cycles and different grades of hemolytic plasma could affect amount of cfDNA and gDNA. Samples of Study 1 (analysis of influence of freeze–thaw cycles) were pooled aliquoted samples ($n = 89$) and samples of Study 2 (analysis of hemolytic plasma interference) were individual aliquoted samples ($n = 36$). Only samples isolated by the QIAamp Kit showed significant changes in cfDNA yield concerning plasma biobanking conditions: the amounts of cfDNA in fresh plasma samples (2.46 ± 0.61 ng/mL of plasma), in samples after one freeze–thaw cycle (3.15 ± 0.55 ng/mL of plasma) and in samples
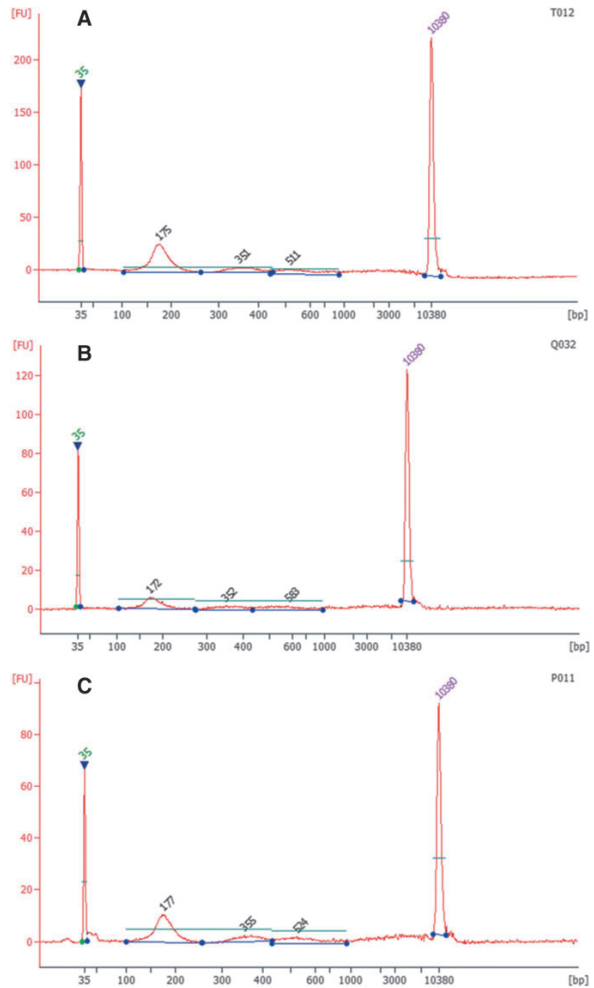
FIG. 2. Plasma cfDNA length distributions depending on kit. The figure shows representative Agilent Bioanalyzer 2100 Electropherograms of cfDNA from pooled plasma samples. **(A)** Isolated with the MagMAX Kit from pooled healthy plasma. **(B)** Isolated with the QIAamp Kit from pooled healthy plasma. **(C)** Isolated with the NextPrep Kit from pooled healthy plasma. The cfDNA length peaks around 180, 350p, and 500 bp were considered as mono-, di-, and trinucleosomal cfDNA fragments, respectively.

TABLE 2. AMOUNT OF CELL-FREE DNA AND GENOMIC DNA ISOLATED BY DIFFERENT ISOLATION KITS

	Qubit (ng/mL of plasma)			Agilent (ng/mL of plasma)		
	NextPrep-Mag (n=41)	MagMAX (n=42)	QIAamp (n=42)	NextPrep-Mag (n=41)	MagMAX (n=42)	QIAamp (n=42)
cfDNA	5.74±1.57	6.06±2.08	9.49±4.02	4.15±1.36	4.16±1.44	3.87±1.85
gDNA	—	—	—	0.53±0.44	0.81±0.56	0.90±0.69
CV of cfDNA yield (%)	12.32	11.68	14.91	30.05	14.12	26.30

cfDNA, cell-free DNA; CV, coefficient of variation; gDNA, genomic DNA.

4

115
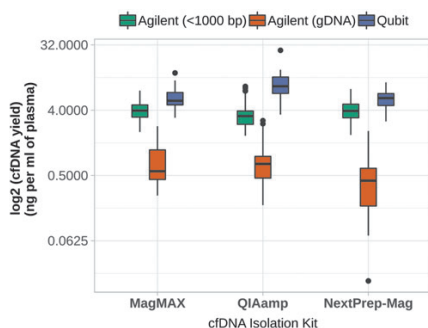
FIG. 3. Amount of cfDNA isolated by different kits and quantified by different methods. Yield of cfDNA in all plasma samples (Study 1 and Study 2 samples) analyzed in the study. Based on Qubit quantification, a significantly higher amount of cfDNA was isolated with QIAamp. Based on Agilent quantification, no statistical significance was determined between the cfDNA yields of different isolation kits. The amount of gDNA was higher in isolations by QIAamp and MagMAX than by NextPrep. *Boxes* represent interquartile range, from the *bottom*: 25th percentile, 50th percentile, and 75th percentile. *Whiskers*—largest and smallest value within 1.5 times interquartile range. *Dots* represent outside values—>1.5 times and <3 times interquartile range beyond either end of the box.

after two freeze–thaw cycles (3.11 ± 0.63 ng/mL of plasma) showed a significant increase in cfDNA yield (*T*-test, $p = 0.016$ and $p = 0.031$, respectively) (Fig. 4).

No significant differences in cfDNA yield after freeze–thaw cycles were observed in samples isolated by MagMAX

and NextPrep-Mag kits. Moreover, no significant differences were seen between the amounts of gDNA from different plasma storage conditions and isolation kits. Comparing nonhemolyzed plasma with moderately or severely hemolyzed plasma samples, no significant differences were determined in the cfDNA amounts in samples isolated by three different kits. However, significant differences were determined for the amounts of gDNA in samples isolated by MagMAX between nonhemolyzed samples and samples with severe hemolysis (1.34 ± 0.18 ng/mL vs. 2.12 ± 0.23 ng/mL of plasma; *T*-test, $p = 0.002$) (Fig. 5).

Amount of gDNA and miRNA in cfDNA samples depends on isolation kit

The amount of co-occurring gDNA and miRNA in the cfDNA isolates is summarized in Figures 3 and 6, respectively. The amount of gDNA was significantly higher for the QIAamp (0.90 ± 0.69 ng/mL of plasma) and MagMAX (0.81 ± 0.56 ng/mL of plasma) Isolation Kits compared with the NextPrep-Mag (0.53 ± 0.45 ng/mL of plasma) Isolation Kit (Mann–Whitney *U* test, $p = 0.001$, $p = 0.004$, respectively) (Fig. 3). miR-223 quantification analysis showed significantly higher amounts of miRNA in samples isolated by QIAamp and MagMAX ($2 \times 10^{-3} \pm 9 \times 10^{-4}$ and $7 \times 10^{-3} \pm 7 \times 10^{-3}$ pg/1 mL of plasma; *T*-test, $p = 0.003$ and $p = 0.027$, respectively) compared with samples isolated by MagMAX ($1 \times 10^{-4} \pm 8 \times 10^{-5}$ pg/1 mL of plasma) (Fig. 6).

Distribution of cfDNA fragment lengths varies depending on isolation kit and also could be affected by plasma storage conditions

The cfDNA length distributions depending on kit are shown in Figure 2. These length distributions show three peaks that we interpreted to be mono-, di-, and
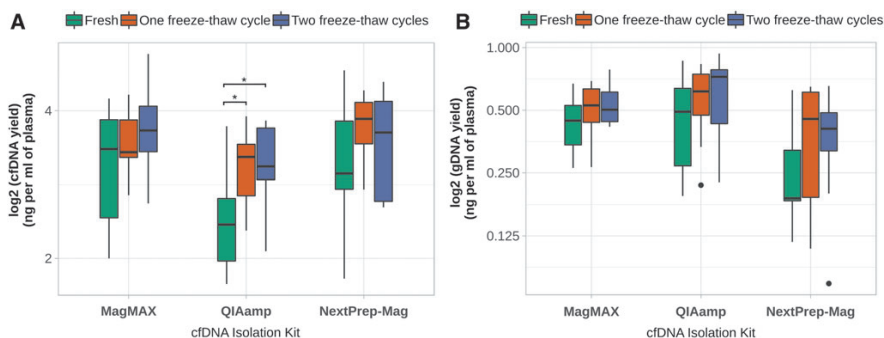

FIG. 4. Distribution of cfDNA amount according to different plasma storage conditions. **(A)** Yield of cfDNA and **(B)** gDNA in plasma samples under different plasma storage conditions (fresh, one, and two freeze–thaw cycles). An increase in cfDNA yield was observed in samples isolated by QIAamp after one and two freeze–thaw cycles. No significant differences in amount of gDNA were determined when comparing different plasma storage conditions and isolation kits. All yields are based on Agilent quantification. *Boxes* represent interquartile range, from the *bottom*: 25th percentile, 50th percentile, and 75th percentile. *Whiskers*—largest and smallest value within 1.5 times interquartile range. *Dots* represent outside values—>1.5 times and <3 times interquartile range beyond either end of the box. (*$p < 0.05$).
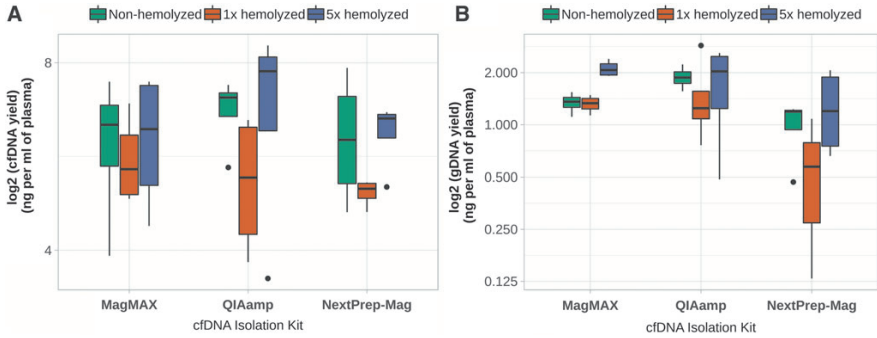
FIG. 5. Amount of cfDNA and gDNA in nonhemolyzed and hemolyzed samples. **(A)** Yield of cfDNA and **(B)** gDNA in nonhemolyzed samples and samples with moderate and severe hemolysis ($1\times$ and $5\times$ hemolyzed, respectively). No significant differences were determined between the amounts of cfDNA in samples isolated by different isolation kits. A significant increase in gDNA yield was determined in samples isolated by MagMAX between nonhemolyzed samples and samples with severe hemolysis. All yields are based on Agilent quantification. *Boxes* represent interquartile range, from the *bottom*: 25th percentile, 50th percentile, and 75th percentile. *Whiskers*—largest and smallest value within 1.5 times interquartile range. *Dots* represent outside values—>1.5 times and <3 times interquartile range beyond either end of the box.

trinucleosomal cfDNA fragments. Comparing performance of the different isolation kits, fresh plasma samples were analyzed ($n = 29$); while comparing plasma storage conditions—fresh plasma samples and samples after one or two freeze–thaw cycles were analyzed ($n = 89$). A higher amount of mononucleosomal fragments was determined in samples
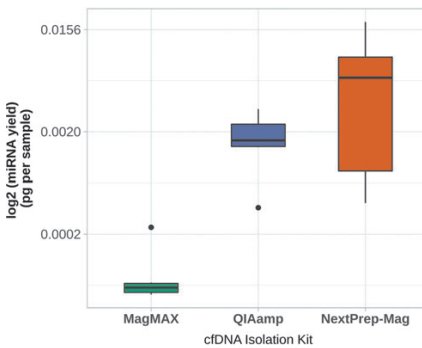
FIG. 6. Amount of miRNA in plasma samples isolated by different kits. Yield of miR-223 in pooled plasma samples ($n = 3$). Significantly higher amount of miRNA in samples isolated by QIAamp and NextPrep was determined compared with samples isolated by MagMAX. Yields based on qPCR. *Boxes* represent interquartile range, from the *bottom*: 25th percentile, 50th percentile, and 75th percentile. *Whiskers*—largest and smallest value within 1.5 times interquartile range. *Dots* represent outside values—>1.5 times and <3 times interquartile range beyond either end of the box. qPCR, quantitative polymerase chain reaction.

isolated by NextPrep-Mag (2.78 ± 0.97 ng/mL of plasma) and MagMAX (2.89 ± 1.04 ng/mL of plasma) Isolation Kits than in samples isolated by QIAamp (2.09 ± 1.07 ng/mL of plasma; Mann–Whitney U test, $p = 0.0003$ and $p = 8.22 \times 10^{-5}$, respectively).

On the other hand, samples isolated by QIAamp showed a higher amount of dinucleosomal fragments (0.94 ± 0.46 ng/mL of plasma) compared with NextPrep-Mag (0.67 ± 0.28 ng/mL of plasma) and MagMAX (0.66 ± 0.31 ng/mL of plasma) (Mann–Whitney U test, $p = 0.002$ and $p = 0.001$, respectively). Samples isolated by QIAamp also showed a higher amount of trinucleosomal fragments compared with samples isolated by MagMAX (0.85 ± 0.52 ng/mL of plasma and 0.59 ± 0.23 ng/mL of plasma respectively; Mann–Whitney U test, $p = 0.008$) (Fig. 7).

The amounts of mono-, di-, and trinucleosomal cfDNA depending on the number of freeze–thaw cycles are shown in Figure 8. A slight increase of mononucleosomal fragments in samples isolated by QIAamp after one freeze–thaw cycle was observed (1.31 ± 0.42 and 1.70 ± 0.38 ng/mL of plasma, fresh and once frozen respectively; T-test, $p = 0.044$). A slight increase in trinucleosomal fragments in samples isolated by QIAamp after two freeze–thaw cycles was determined (0.51 ± 0.22 and 0.72 ± 0.15 ng/mL of plasma, fresh and twice frozen respectively; T-test, $p = 0.029$). A slight increase in mononucleosomal fragments was observed in samples isolated by MagMAX after two freeze–thaw cycles (2.11 ± 0.59 and 2.77 ± 0.60 ng/mL of plasma, fresh and twice frozen respectively; T-test, $p = 0.023$) (Fig. 8).

Discussion

Recent studies of cancer biomarkers mainly focus on minimally invasive tests, such as liquid biopsies, which could be important for regular monitoring, disease management (relapse, metastasis), and evaluation of resistance to therapy. The minimally invasive nature of plasma
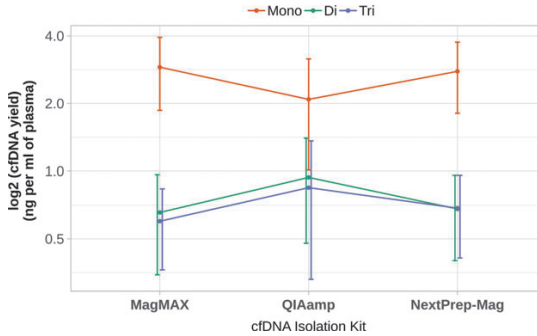
FIG. 7. Distribution of cfDNA fragments isolated with different isolation kits. Distribution of cfDNA fragments in all plasma samples analyzed in the study and isolated by three different kits. Higher amount of mononucleosomal fragments was determined in samples isolated by the NextPrep-Mag and MagMAX Isolation Kits compared with QIAamp. Higher amount of di- and trinucleosomal fragments in samples isolated by QIAamp than by other isolation kits. Yields based on Agilent quantification.

sampling makes cfDNA a suitable clinical analyte. However, isolation of cfDNA has some procedural features that should be taken into account as they can affect downstream applications.

In our study, we evaluated the impact of the quantification method and found that the amount of cfDNA quantified by Qubit and Agilent differed significantly. cRNA is present in QIAamp isolations, which could interfere with the signals as also stated in the handbook presented by the manufacturer. Spectrometry and fluorescence-based methods are less specific and in the special case when cRNA (QIAamp isolation method) is used, the quantification signal may be increased nonspecifically.[17] Real-time PCR is currently recommended as the gold standard for DNA quantification in biobanks. However, this method is expensive and time consuming.

On the other hand, the microcapillary electrophoresis-based method has many advantages: it is less time consuming, provides information on the distribution of cfDNA fragment lengths and on the presence of gDNA, and it is specific to DNA and sensitive. Indeed, cRNA was not observed to impact the cfDNA quantification when using Agilent (data not shown).

Previous studies widely reported the advantages of the QIAamp Isolation Kit, and Qiagen remains one of the leaders in this field.[10,11,13,18] It is important to note that in many of those cases cfDNA yield was quantified by fluorescence-based methods, which could lead to inaccurate conclusions. Similarly, our results showed a significantly greater amount of cfDNA isolated by this QIAamp Kit only when quantified by Qubit fluorometer, and in this case cRNA interference could not be ruled out. In terms of CVs of pools of plasma samples quantified by Qubit and Agilent, MagMAX showed the lowest CVs of 11.68% and 14.12%, respectively, suggesting that this kit may enable the best technical reproducibility.

Second, significant changes in the cfDNA amount were observed in samples isolated by the QIAamp Isolation Kit after additional plasma freeze–thaw cycles, while, interestingly, no increase of gDNA was determined. In our study we considered fragments longer than 1000 bp to be gDNA. The Agilent High-Sensitivity DNA Kit size range is 50–7000 bp, and larger fragments were not analyzed, therefore results on gDNA presence should be interpreted with caution. Third, no significant changes in isolated cfDNA or gDNA amounts were observed in hemolyzed plasma samples,
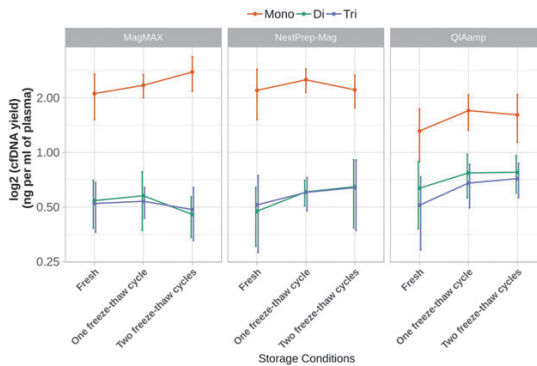

FIG. 8. Distribution of cfDNA fragments isolated with different isolation kits after different plasma storage conditions. Distribution of cfDNA fragments in plasma samples under different plasma storage conditions (fresh, one, and two freeze–thaw cycles). Slight increase of mononucleosomal fragments in samples isolated by QIAamp after one freeze–thaw cycle was observed, and slight increase in trinucleosomal fragments in samples isolated by QIAamp after two freeze–thaw cycles was determined. Also, slight increase in mononucleosomal fragments was observed in samples isolated by MagMAX after two freeze–thaw cycles compared with fresh plasma samples.

except for an increase of gDNA when using the MagMAX Isolation Kit.

Fourth, we analyzed the yield of larger DNA fragments and miRNA in cfDNA samples. QIAamp and MagMAX isolations had a higher gDNA background compared with NextPrep-Mag. Although two of isolation kits (MagMAX and NextPrep-Mag) were not designed for miRNA isolation, we were able to show that circulating plasma miRNA is present in cfDNA samples. Absolute miR-223 quantification showed a significantly higher amount of miRNA isolated by QIAamp and NextPrep-Mag compared with MagMAX, suggesting possible application of both kits for analysis of both circulating nucleic acids (DNA and RNA) and even multiomics approaches.

Finally, we analyzed the recovery rates of mono-, di-, and trinucleosomal fragments, which are important for planning sequencing strategies. We evaluated these recovery rates for each isolation kit and for different plasma biobanking conditions. Our results revealed that the kits tested showed significantly different recovery. Greater amounts of small fragments were isolated by MagMAX and NextPrep-Mag, and greater amounts of bigger fragments were isolated by QIAamp.

Storage conditions also had an impact on recovery: an increase in mono- and trinucleosomal fragments was seen in samples isolated by QIAamp after one freeze–thaw cycle, and a slight increase in mononucleosomal fragments was seen in samples isolated by MagMAX after two freeze–thaw cycles. The increase in mononucleosomal fragments could partially be interpreted as the degradation of longer fragments after freeze–thaw cycles. The increase in trinucleosomal fragments may be due to the rupturing of cellular debris and therefore the recovery of longer fragments. Considering hybridization capture-based sequencing, length of the cfDNA fragments play an important role. Modifications of sequencing library preparation and especially DNA shearing step should be taken into account as well.

This study has some potential limitations. First, small sample size; in the analysis of different plasma pools different CVs were seen, showing high data variability. Increasing the number of samples would also give greater statistical power to detect differences. Second, the amounts of cfDNA in healthy individuals are low, therefore further studies with cancer patients may show a wider range of factors that could impact cfDNA yield, for example, saturation of beads or columns. Finally, qPCR or digital droplet PCR analysis, which are currently recommended for cfDNA quantification was not performed to compare both Qubit and Agilent quantification methods.

In conclusion, the quantification method has a major role in cfDNA analysis and further applications, all isolation methods had different strengths and weaknesses that would appeal to different users with different requirements:(1) The NextPrep-Mag Isolation Kit is offered for a lower price (cheapest for isolations of 1 mL samples), has a short and easy protocol (processing time around 45 minutes), recovers sufficient miRNA; (ii) MagMAX showed great results with amount of cfDNA, offers wide range of plasma volume input, however, is more expensive for isolations of 5 mL samples and could potentially be affected by hemolytic plasma. Moreover, proteinase K is not provided. Both bead-based methods require additional equipment (magnetic rack); (iii) QIAamp has a strong economical advantage for samples with greater volume (5 mL), and could be used for circulating miRNA analysis. However, it contains poly-A cRNA, which should be taken into account when choosing the quantification method. Moreover, it is a vacuum-based method requiring expensive additional equipment (vacuum system).

Author Disclosure Statement

No conflicting financial interests exist.

References

1. Thierry AR, El Messaoudi S, Gahan PB, et al. Origins, structures, and functions of circulating DNA in oncology. Cancer Metastasis Rev 2016;35:347–376.
2. Kornberg RD, Lorch Y. Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. Cell 1999; 98:285–294.
3. Wan JCM, Massie C, Garcia-Corbacho J, et al. Liquid biopsies come of age: Towards implementation of circulating tumour DNA. Nat Rev Cancer 2017;17:223–238.
4. Kamat AA, Bischoff FZ, Dang D, et al. Circulating cell-free DNA: A novel biomarker for response to therapy in ovarian carcinoma. Cancer Biol Ther 2006;5:1369–1374.
5. Bettegowda C, Sausen M, Leary RJ, et al. Detection of circulating tumor DNA in early- and late-stage human malignancies. Sci Transl Med 2014;6:224ra24.
6. Sorenson GD, Pribish DM, Valone FH, et al. Soluble normal and mutated DNA sequences from single-copy genes in human blood. Cancer Epidemiol Biomarkers Prev 1994;3:67–71.
7. Volik S, Alcaide M, Morin RD, et al. Cell-free DNA (cfDNA): Clinical significance and utility in cancer shaped by emerging technologies. Mol Cancer Res 2016;14(10):898–908.
8. Markus H, Contente-Cuomo T, Liang WS, et al. Evaluation of pre-analytical factors affecting plasma DNA analysis. bioRxiv 2017;126839.
9. van Ginkel JH, van den Broek DA, van Kuik J, et al. Preanalytical blood sample workup for cell-free DNA analysis using Droplet Digital PCR for future molecular cancer diagnostics. Cancer Med 2017;6(10):2297–2307.
10. Raymond CK, Hernandez J, Karr R, et al. Collection of cell-free DNA for genomic analysis of solid tumors in a clinical laboratory setting. PLoS One 2017;12:e0176241.
11. Sherwood JL, Corcoran C, Brown H,et al. Optimised pre-analytical methods improve KRAS mutation detection in circulating tumour DNA (ctDNA) from patients with non-small cell lung cancer (NSCLC). PLoS One 2016;11: e0150197.
12. El Messaoudi S, Rolet F, Mouliere F, et al. Circulating cell free DNA: Preanalytical considerations. Clin Chim Acta 2013;424:222–230.

13. Pérez-Barrios C, Nieto-Alcolado I, Torrente M, et al. Comparison of methods for circulating cell-free DNA isolation using blood from cancer patients: Impact on biomarker testing. Transl lung cancer Res 2016;5:665–672.

14. Fong SL, Zhang JT, Lim CK, et al. Comparison of 7 methods for extracting cell-free DNA from serum samples of colorectal cancer patients. Clin Chem 2009;55:587–589.

15. Koseoglu M, Hur A, Atay A, et al. Effects of hemolysis interference on routine biochemistry parameters. Biochem Med 2011;79–85.

16. Dolci A, Panteghini M. Harmonization of automated hemolysis index assessment and use: Is it possible? Clin Chim Acta 2014;432:38–43.

17. Bronkhorst AJ, Aucamp J, Pretorius PJ. Cell-free DNA: Preanalytical variables. Clin Chim Acta 2015;450:243–253.

18. Medina Diaz I, Nocon A, Mehnert DH, et al. Performance of Streck cfDNA blood collection tubes for liquid biopsy testing. PLoS One 2016;11:e0166354.

Address correspondence to:
Jurgita Skieceviciene, PhD
Institute for Digestive Research
Lithuanian University of Health Sciences
Eiveniu Street 4
50161 Kaunas
Lithuania

E-mail: jurgita.skieceviciene@lsmuni.lt

A5

Title: Quantifying sequencing error and effective sequencing depth of liquid biopsy NGS with UMI error correction

Authors: Frank, Malene Støchkel; Fuß, Janina; Steiert, Tim Alexander; Streleckienė, Greta; Gehl, Julie; Forster, Michael

BioTechniques 70, 4 (2021)

Benchmark

Quantifying sequencing error and effective sequencing depth of liquid biopsy NGS with UMI error correction

Malene Støchkel Frank[‡,1,2] , Janina Fuß[‡,3] , Tim Alexander Steiert[3] , Greta Streleckiene[4] , Julie Gehl[1,2] & Michael Forster*[,3]

[1]Department of Clinical Oncology & Palliative Care, Zealand University Hospital, Køge, Denmark; [2]Department of Clinical Medicine, Faculty of Health & Medical Sciences, University of Copenhagen, Copenhagen, Denmark; [3]Institute of Clinical Molecular Biology, Christian Albrechts University Kiel, Kiel 24105, Germany; [4]Institute for Digestive Research, Lithuanian University of Health Sciences, Kaunas 50161, Lithuania; *Author for correspondence: m.forster@ikmb.uni-kiel.de; [‡]Authors contributed equally

ABSTRACT
Liquid biopsies are a minimally invasive method to diagnose and longitudinally monitor tumor mutations in patients when tissue biopsies are difficult (e.g., in lung cancer). The percentage of cell-free tumor DNA in blood plasma ranges from more than 65% to 0.1% or lower. To reliably diagnose tumor mutations at 0.1%, there are two options: unrealistically large volumes of patient blood or library preparation and sequencing depth optimized to low-input DNA. Here, we assess two library preparation methods and analysis workflows to determine feasibility and reliability based on standards with known allelic frequency (0 and 0.13% in *PIK3CA*). However, the implementation for patients is still costly and requires elaborate setups.

GRAPHICAL ABSTRACT

122
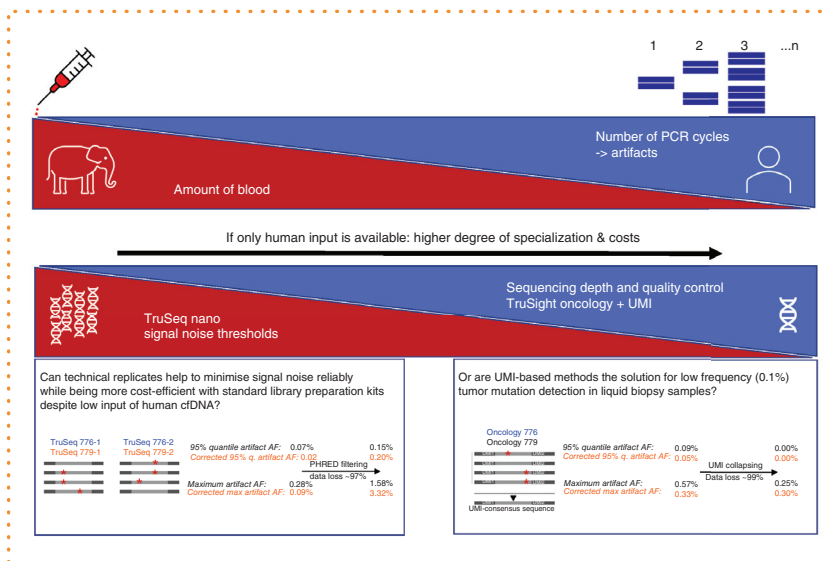
Benchmark

METHOD SUMMARY

Two widely used Illumina library preparation kits were benchmarked for next-generation sequencing of cell-free tumor DNA: one kit without unique molecular identifiers (UMIs) but in technical duplicates and one kit with UMIs. Two reference cell-free DNA samples were used with 0 and 0.13% tumor allele frequency, respectively. Targeted sequencing was performed to 50,000× average depth. Illumina's UMI Error Correction Local App was used for aligning and collapsing UMI sequences. Signal noise in UMI- and non-UMI libraries was compared, and effective sequencing depth loss due to the bioinformatic processing was quantified to allow for estimating the required sequencing depth.

Cancer is the second leading cause of death worldwide, with lung cancer being the cancer entity causing the highest number of fatalities [1]. Approximately 40–50% of patients with non-small-cell lung cancer (NSCLC) are diagnosed with advanced or metastatic disease and are not candidates for curative therapy but only for life-prolonging and palliative treatment. After the exhaustion of standard treatments, some patients may be discussed by molecular tumor boards and included into clinical trials based on their tumor mutations. These mutations are usually revealed by a tissue biopsy, but liquid biopsy is gaining interest because of its ability to reveal tumor mutations in a minimally invasive manner. Furthermore, liquid biopsy can potentially predict treatment response and detect minimal residual disease after cancer surgery. The percentage of cell-free tumor DNA in blood plasma ranges from more than 65% in advanced patients [2] to 0.1% or lower in lung cancer patients with less advanced disease [3]. Feasibility, sensitivity/specificity and cost-effectiveness are crucial factors in clinical implementation. Thus, an overall aim must be to supply reliable data to the individual lung cancer patient and, at the same time, to set up the analyses to be simple and inexpensive enough that services may be applied broadly. In this article, we sought to explore these factors for two methods of library preparation and analysis to determine a cost-efficient and robust workflow and to determine the thresholds of artifacts that would confound the detection of low percentages of tumor. We used the Multiplex I cfDNA Reference Standard Set (Horizon Discovery Ltd., Cambridge, UK): the 100% Multiplex I Wild Type cfDNA Reference Standard (Cat #HD776) and the 0.1% Multiplex I cfDNA Reference Standard (Cat #HD779) with mutated *PIK3CA* p.E545K (chr3: g.178936091G>A in the hg19 genome version) of 0.13% allele frequency according to the supplier. From each DNA, relatively inexpensive libraries without UMIs were prepared in duplicates (TruSeq Nano, Cat #FC-121-4001, Illumina Inc., CA, USA) using a modified protocol for low input and small fragment sizes (160 bp fragments according to Horizon Discovery). In short, modifications to the protocol were the omission of the shearing step due to already fragmented DNA and therefore starting at the end repair step as the entry point with 50 ng of cell-free DNA (cfDNA), SPRIselect bead (SPB) cleanup without size selection (2× vol SPB) and 12 cycles of PCR instead of eight. From 40 ng of each standard, libraries with unique molecular identifier (UMIs) for cfDNA (TruSight Oncology UMI Reagents, Cat #20024586, Illumina Inc.) were prepared according to protocol, and 500 ng of all six libraries were pooled in equimolar amounts. Targeted capture was done according to the protocol for xGen hybridization capture of DNA libraries (Integrated DNA Technologies, Inc., Coralville, IA, USA), with custom oligos for coding regions of *PIK3CA* supplied by Illumina, and with 10 PCR cycles. The captured pool was sequenced with an Illumina NextSeq 500 using 2 × 150 bp paired-end reads. We aimed for the raw sequencing depth in the target region to be approximately 50,000×. The sequences were aligned to hg19 using bwa [4] and sorted with samtools [5]. For the settings of all bioinformatic tools used, see Supplementary Table 1. Library type-specific bioinformatic filtering was performed as follows: The TruSight Oncology UMI sequences were aligned and collapsed to consensus sequences using the UMI Error Correction Local App (Illumina Inc.). This software package identifies 'families' of sequence duplicates that are aligned to the same genomic position and that have the same UMI (see Figure 1) and replaces each 'family' of duplicates with a single consensus sequence, resulting in a 'stitched' bam file of aligned, error-corrected consensus sequences, thus reducing the impact of randomly occurring PCR artifacts on the final alignment and variant calling as well as reducing potential sequencing artifacts. The UMIs are unique sequences of seven-nucleotides length that are added to each end of a cfDNA fragment during library preparation before PCR is performed. From the raw sequencing files in binary base call (BCL) format, the UMI Error Correction Local App performs base calling, adds the UMI sequences into the sequence names, performs sequence alignment using bwa [4] and samtools [5] and collapses families of sequences that were aligned to the same genomic position and that have the same UMI (see Figure 1) into an error-corrected consensus sequence. This is an automatic process in which two default parameters can be changed: the alignment quality threshold (default: 0) and the minimal read family threshold (default: 2). We used the manufacturer's settings, according to standard laboratory practices and International Organization of Standardization recommendations.

For the TruSeq Nano sequences (without UMIs), stringent bioinformatic quality filtering was performed on the raw sequences before alignment to assess whether the signal noise could be reduced compared with standard filtering, that is, to remove sequence instrument-based artifacts. This stringent filtering was performed by eliminating all paired-end sequences that did not meet a phred-scaled base quality of 32 for each base, using fastp [6].

To test whether the combined analysis of the technical replicates can reduce signal noise in the TruSeq Nano library duplicates, an artificial dataset was created. We thus pooled randomly extracted sequence read-pairs from each of the two TruSeq Nano library duplicates' sequence data files into a single sequence data set. For better comparability between all sequence data sets, we normalized
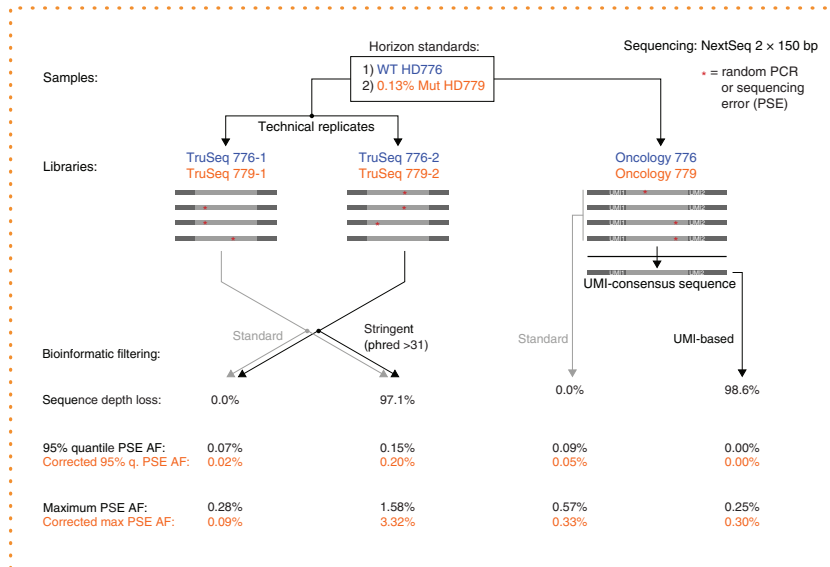
Figure 1. Schematic representation of analysis workflow. Horizon cell-free DNA reference standards HD776 (WT) and HD779 (0.13% known allelic frequency of g.178936091G>A, hg19) were both prepared with Illumina TruSeq Nano and TruSight Oncology library preparation kits. For TruSight Oncology (with unique molecular identifier [UMIs]) error-corrected consensus sequences were generated from read families and compared with standard filtering. For TruSeq Nano (without UMIs), standard filtering was compared with stringent filtering and to the TruSight Oncology results. The quality metrics for the comparisons were the 95% quantile of the PCR and sequencing errors' (PSE) allelic frequency (AF) and maximum PSE AF. The PSE AFs were corrected (and rounded to two digits) on the basis of the HD779 standard with its 0.13% mutation allelic frequency.

the pooled library sequence sets (TruSeq 776-Norm, TruSeq 779-Norm) to have approximately identical sequencing depths. Normalization consisted of random sequence data extraction, alignment, computation of sequence depth in the target region (specifically, at the *PIK3CA* mutation) and iterative repetition of these steps until the sequencing depth in the target region in the normalized TruSeq Nano libraries was comparable with the respective TruSight Oncology libraries. In theory, this bioinformatic process should reduce the randomly appearing artifacts introduced by the significant amount of PCR cycles in the library preparation because those should appear randomly and independently in technical replicates but should not reduce the artifacts arising from the sequencing instrument.

BAM files of the sequences were deposited at the European Nucleotide Archive under Accession http://www.ebi.ac.uk/ena/data/view/PRJEB39899. GenSearchNGS [7] (PhenoSystems S.A., Braine le Chateau, Belgium) was used to analyze the bam files for low allele frequency variants in joint-variant-calling mode, with and without the option 'ignore duplicate reads'. The 'ignore duplicate reads' option was not used for the analysis of the UMI-based error corrected 'stitched' bam files because the redundant sequences were already collapsed into error-corrected consensus sequences by the UMI Error Correction Local App. The resulting two multisample variant lists (with and without duplicate reads, i.e., redundant sequences) were exported in csv format. Allele frequencies were recomputed from the 'coverage' and 'reads' columns. Lower coverage genomic positions (less than 500× in the alignments) were filtered. Box plots and density plots of low-allele-frequency PCR and sequencing errors (signal noise) for each library type were plotted in R. The allelic frequency of the signal noise was corrected as follows using the known allelic frequency of the mutation in HD779 versus the sequenced allelic frequency values:

$$AF_{corrected} = \frac{AF_{uncorrected} \times 0.13\%}{AF_{at\ E545\ mutation}}$$

Benchmark

Type	Library	Filtering[†]	Mean coverage[‡]	N[§]	AF[¶] 95% quantile	AF[¶] max
Non-UMI	TruSeq 776-1	–	55472	9902	**0.07%**	**0.28%**
		phred >31	2160	7108	0.10%	0.79%
	TruSeq 776-2	–	52573	9902	0.07%	0.25%
		phred >31	1416	7108	0.12%	1.58%
	TruSeq 776-Norm	–	51340	9902	0.07%	0.25%
	TruSeq 779-1	–	47552	9902	0.07%	0.20%
		phred >31	789	7108	0.15%	0.77%
	TruSeq 779-2	–	57491	9902	0.07%	0.25%
		phred >31	1948	7108	0.10%	0.80%
	TruSeq 779-Norm	–	61746	9902	0.07%	0.22%
UMI	Oncology 776	–	36786	9963	0.09%	0.51%
		Consensus	602	7141	0.00%	0.25%
	Oncology 779	–	48924	9963	0.09%	0.57%
		Consensus	**670**	**7141**	**0.00%**	**0.25%**

[†]Filtering: phred >31 considered only read-pairs with base quality >31 for all bases in the read-pair. Consensus filtering with the Illumina UMI app collapsed duplicate reads with identical UMIs to a consensus read.
[‡]Mean coverage: considered the entire coding region of *PIK3CA* but counted only bases with at least phred >19.
[§]N: number of observations (genomic positions in *PIK3CA* with sequencing depth at least 500× in all libraries of the same type).
[¶]AF: Allele frequency of the PCR and sequencing errors, the computation of which considered only those genomic coordinates with a minimal sequencing depth of 500× in libraries of the same group (TruSeq unfiltered, TruSeq filtered, Oncology unfiltered, Oncology filtered).

The main results can be summarized as follows (Figure 1).

The allele frequency of PCR and sequencing errors (signal noise) was reproducibly lower than 0.07% for the four non-UMI libraries in 95% of the considered genomic positions (Table 1) and exactly 0.00% for the two UMI libraries in 95% of positions. However, numerous PCR and sequencing error outliers were detected with higher allele frequencies of up to 0.28% in the four non-UMI libraries (Figure 2A & Table 1) and up to 0.25% in the two UMI libraries with consensus filtering (Figure 2D & Table 1), after ensuring that the effective sequencing depth at each considered position was at least 500 reads. The PCR and sequencing errors' allele frequencies were higher if lower effective sequencing depths were permitted. For the non-UMI libraries, stringent bioinformatic filtering reduced sequencing depths and increased the allele frequency of the PCR and sequencing errors. On the other hand, for the UMI libraries, bioinformatic error correction lowers the allele frequency of the PCR and sequencing errors, despite reducing the effective sequencing depth (Table 1). Of particular note for those who are responsible for the specification and the costs of experiments: in our combination of Illumina and IDT protocols, the mean effective sequencing depth in the *PIK3CA* target region was reduced 60-fold by the bioinformatic error-processing of UMI libraries.

The known 0.13% *PIK3CA* p.E545K mutation can be used to calibrate the allelic frequencies obtained from the sequencing reads (Supplementary Table 2) and then used to correct the allelic frequency of the PCR and sequencing errors. Table 2 shows the corrected summary values. For the non-UMI libraries, the calibrated signal noise is up to threefold lower than before calibration, indicating high levels of signal noise – probably PCR duplicates – at the *PIK3CA* mutation site and therefore at all sites. For the UMI libraries after bioinformatic consensus filtering, the signal noise is nearly unchanged after calibration. This is confirmed by the more accurate tumor allele frequency of 0.1% for the UMI-based error corrected sequences at the *PIK3CA* site compared with 0.4% from the libraries without UMIs (Supplementary Table 2).

There are also three secondary results that may be of interest to a laboratory: first, when UMI-based error correction bioinformatics are used, the conventional quality criterion of a well-balanced ratio of forward–reverse sequences is not applicable (Supplementary Figure 1). This is a purely bioinformatic result after the UMI-based error correction because the original sequences are well balanced before error correction. Second, in the non-UMI libraries, bioinformatic removal of redundant sequences (deduplication) was detrimental for the signal-to-noise ratio because it reproducibly amplified the signal noise (Supplementary Table 3) but reduced the allele frequency of the true mutation (Supplementary Table 2). Third, the PCR and sequencing errors' allele frequencies were not improved in our normalization sub-experiment (TruSeq 776-Norm, TruSeq 779-Norm), see Table 1 & Supplementary Table 3, suggesting that PCR error was not the dominant contributor to the total sequencing error.

In conclusion, the mean loss of effective sequence depth in the UMI libraries after bioinformatic error correction is on the order of 98.5%, requiring 60-fold deeper sequencing than for conventional non-UMI libraries. This problem is due to the small amounts of cfDNA that are available from a human blood sample and the resulting 25 cycles of PCR that were needed in the UMI-based protocol. By comparison, the original report by Schmitt and colleagues [8] on UMI-based error correction demonstrated its principle using 3 μg of input
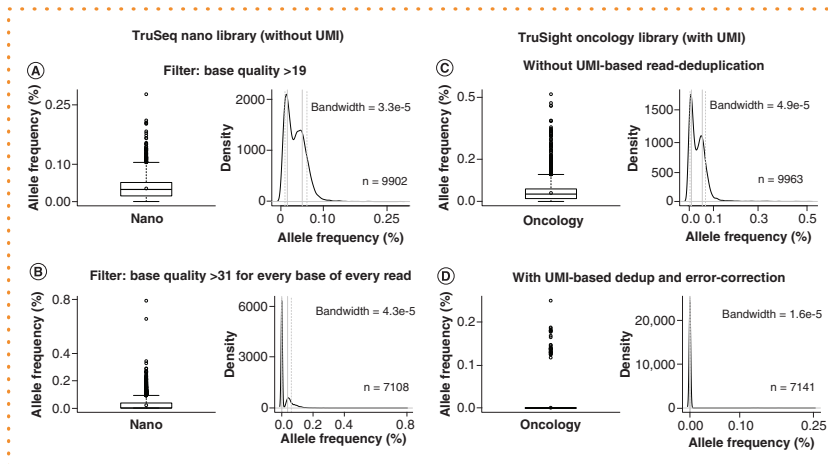
125

Figure 2. Allele frequency distributions of noncalibrated PCR and sequencing errors (signal noise) depending on kit and bioinformatic processing. The plots show the signal-noise allele frequency (AF) ranges for the wild-type Horizon standard HD776, from targeted sequencing of the *PIK3CAa* coding regions. The AFs are the 'as-is' AFs, that is, without the HD779 mutation-based correction/calibration. The density plots include the 12.5%, 25%, 75% and 87.5% quantiles. **(A)** Non–unique molecular identifier (UMI) library type TruSeq Nano, where only base calls with base quality >19 were considered. Without calibration, the max AF is 0.28%; with calibration it is only 0.09%. The density plot shows a bimodal distribution of allele frequencies, which might be explained by a large number of high-quality base calls with small AFs (left peak) and a slightly smaller number of lower quality base calls with higher AFs (right peak), because the right peak nearly disappears when stringent base quality filtering is applied (see panel B). The right peak (higher AFs) is not due to sequence duplication because a similar bimodal distribution can be seen after filtering duplicates or redundant sequences (data not shown). **(B)** Non-UMI library type TruSeq Nano, where only highest quality read-pairs (all base qualities >31) were used for alignment and variant calling. In the density plot, the majority of AFs are exactly zero, and a small peak can be seen at 0.05% AF. Of note, the max AF has risen from 0.28% to 0.8%, which is explained by the lower sequencing depth after stringent filtering. **(C)** UMI-library type TruSight Oncology, where conventional alignment was performed without utilizing the UMIs for error correction or read deduplication. **(D)** UMI-library type TruSight Oncology, where the Illumina app was used for UMI-based deduplication and error correction.

DNA and only 18–20 PCR cycles. We would obtain 3 μg of cfDNA from a hypothetical 1-l blood draw, which would have detrimental health effects on every cancer patient except for the extremely rare elephant with cancer. A 10-ml blood draw from a human patient will result in approximately 4–5 ml of plasma and 40–50 ng of cfDNA. On the other side, Illumina sequencers need a certain loading concentration of library, and a high number of PCR cycles is required in the library preparation of low-input samples. Therefore, the highly duplicated cfDNA molecules will be collapsed to a small fraction after bioinformatic error correction. According to the product information, an increase in sequencing depth will probably not greatly increase the sequence depth of UMI-collapsed consensus reads. Special care should be given to optimization processes in the lab, especially the recovery in the cleanup steps, as well as hybridization and enrichment. Non-UMI libraries have inherently higher signal noise and are less accurate with respect to allele frequency quantification. This means that the allele frequencies of mutations detected with non-UMI-libraries should be calibrated with known tumor allele frequencies by including reference DNA samples into each non-UMI NGS run. Regardless of library type, technical replication is recommended to identify whether variants with low allele frequencies are PCR and sequencing errors. Ideally, control DNAs should be included, such as the Horizon cfDNA set used here.

In summary, this article has identified advantages and drawbacks of two methods for detecting low-frequency alleles in liquid biopsies. At present, applications in cancer care are limited due to high cost. To provide real-life, broad clinical benefit to cancer patients, more optimizations that ensure reliable results in a cost-efficient workflow are crucial.

Future perspective

Currently, cell-free DNA-based cancer tests are offered routinely in the USA for privately insured patients; in many other countries, they are available for patients taking part in research studies. Molecular tumor boards (MTBs) are being initiated in cancer hospitals in many

Benchmark

Table 2. Corrected allelic frequency summary of PCR or sequencing errors for libraries without a unique molecular identifier (UMI) versus UMI libraries and bioinformatic nonfiltering versus filtering, after calibration with known 0.13% *PIK3CA* mutation p.E545K (chr3: g.178936091G>A).

Type	Library	Filtering[†]	Mean coverage[‡]	N[§]	AF[¶] 95% quantile	AF[¶] max
Non-UMI	**TruSeq** 776-1	—	55472	9902	**0.02%**	**0.09%**
		phred >31	2160	7108	0.20%	1.65%
	TruSeq 776-2	—	52573	9902	0.02%	0.08%
		phred >31	1416	7108	0.24%	3.32%
	TruSeq 776-Norm	—	51340	9902	0.02%	0.08%
	TruSeq 779-1	—	47552	9902	0.02%	0.06%
		phred >31	789	7108	0.30%	1.62%
	TruSeq 779-2	—	57491	9902	0.02%	0.08%
		phred >31	1948	7108	0.21%	1.67%
	TruSeq 779-Norm	—	61746	9902	0.02%	0.07%
UMI	Oncology 776	—	36786	9963	0.05%	0.30%
		Consensus	602	7141	0.00%	0.30%
	Oncology 779	—	48924	9963	0.05%	0.33%
		Consensus	670	7141	**0.00%**	**0.30%**

[†] Filtering: phred >31 filtering considered only read-pairs with base quality >31 for all bases in the read-pair. Consensus filtering with the Illumina UMI app collapsed duplicate reads with identical UMIs to a consensus read.
[‡] Mean depth: considered the entire coding region of *PIK3CA* but counted only bases with at least phred >19.
[§] N: number of observations (genomic positions in *PIK3CA* with sequencing depth at least 500× in all libraries of the same type).
[¶] AF: Allele frequency of the PCR and sequencing errors, the computation of which considered only those genomic coordinates with a minimal sequencing depth of 500× in libraries of the same group (TruSeq unfiltered, TruSeq filtered, Oncology unfiltered, Oncology filtered). The AF values in this table have been corrected using the known *PIK3CA* AF of 0.13% versus the sequenced AF in the HD779 DNA. The two methods with the smallest PCR and sequencing errors' AFs are highlighted in bold: non-UMI TruSeq libraries without filtering or UMI Oncology libraries with consensus filtering.

countries, following the model established by leading cancer centers in the USA. The MTBs discuss individual patient cases for potential individualized treatment after exhaustion of standard treatments, such as for the choice of a suitable clinical trial. An ever-increasing portfolio of drugs and drug combinations is being tested on cancer patients in clinical trials, some with mutation-based inclusion criteria. Tissue biopsy has been the gold standard for mutation detection, but the clear advantages of liquid biopsy in longitudinal monitoring of resistance mechanisms with the representativeness of all tumor sites will have a major impact in the field of individualized treatment in the coming years. From our past experience, the cost of high-fidelity blood-based cell-free DNA cancer tests is likely to drop to a 10th of the current costs, allowing their widespread use, for example, for monitoring the therapy response of advanced-stage cancer patients and detecting minimal residual disease after cancer surgery. Many of the clinical guidelines will be updated with recommendations for cell-free DNA testing, based on currently ongoing clinical studies that are exploring the clinical utility of these tests. In approximately 10 years, cost–benefit analyses by public health systems and their professional societies are likely to be completed in all EU countries, defining the indications for which patients with a public health insurance will be entitled to receive free high-fidelity cell-free DNA-based cancer tests. It seems most likely that the methods race will be won by UMI-based sequencing and error-correction methods, using a highly automated high-fidelity sequencing platform that will have a hundred times more sequence output than the current largest instruments.

Supplementary data

To view the supplementary data that accompany this paper please visit the journal website at: www.future-science.com/doi/suppl/10.2144/btn-2020-0124

Financial & competing interests disclosure

Open access

References

Papers of special note have been highlighted as: •• of considerable interest

1. American Cancer Society. Cancer facts & figures 2020. American Cancer Society, GA, USA (2020). www.cancer.org/research/cancerfactsstatistics/cancerfactsfigures2020/index

► 2. Hendricks A, Rosenstiel P, Hinz S *et al.* Rapid response of stage IV colorectal cancer with APC/TP53/KRAS mutations to FOLFIRI and bevacizumab combination chemotherapy: a case report of use of liquid biopsy. *BMC Med. Genet.* 21(1), 3 (2020).
   •• Clinical case report demonstrating the clinical utility of cell-free DNA testing in a stage IV cancer patient.

► 3. Abbosh C, Birkbak NJ, Wilson GA *et al.* Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. *Nature* 545(7655), 446−451 (2017).
   •• Lung cancer study report on sequencing low-input amounts of cell-free DNA with non-UMI sequencing libraries in combination with calibration DNA samples.

► 4. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25(14), 1754−1760 (2009).

► 5. Li H, Handsaker B, Wysoker A *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16), 2078−2079 (2009).

6. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34(17), i884−i890 (2018).

7. Wolf B, Kuonen P, Dandekar T, Atlan D. DNAseq workflow in a diagnostic context and an example of a user friendly implementation. *Biomed. Res. Int.* 2015, 403497 (2015).

► 8. Schmitt MW, Kennedy SR, Salk JJ *et al.* Detection of ultra-rare mutations by next-generation sequencing. *Proc. Natl Acad. Sci. USA* 109(36), 14508−14513 (2012).
   •• Report on the theoretical error rates of the duplex sequencing method using UMIs and error correction, demonstrated in a laboratory experiment using a high-input amount of 3 μg DNA and only 18−20 PCR cycles.

A6

Title: Liquid Biopsy in Gastric Cancer: Analysis of Somatic Cancer Tissue Mutations in Plasma Cell- Free DNA for Predicting Disease State and Patient Survival

Authors: Varkalaitė, Greta; Forster, Michael; Franke, Andre; Kupčinskas, Juozas; Skiecevičienė, Jurgita

Clinical and Translational Gastroenterology 12 (2021)

STOMACH

Liquid Biopsy in Gastric Cancer: Analysis of Somatic Cancer Tissue Mutations in Plasma Cell-Free DNA for Predicting Disease State and Patient Survival

Greta Varkalaite, MSc[1], Michael Forster, PhD[2], Andre Franke, PhD[2], Juozas Kupcinskas, MD, PhD[1,3] and Jurgita Skieceviciene, PhD[1]

INTRODUCTION: **Gastric cancer (GC) diagnosis in late stages and high mortality rates are the main issues that require new noninvasive molecular tools. We aimed to assess somatic mutational profiles in GC tissue and plasma cell-free DNA (cfDNA), evaluate their concordance rate, and analyze the role of multilayer molecular profiling to predict disease state and prognosis.**

METHODS: **Treatment-naive GC patient group (n = 29) was selected. Whole exome sequencing (WES) of GC tissue was performed, and a unique 38-gene panel for deep targeted sequencing of plasma cfDNA was developed. Oncoproteins were measured by enzyme-linked immunosorbent assay, and other variables such as tumor mutational burden and microsatellite instability were evaluated using WES data.**

RESULTS: **The yield of cfDNA was increased 43.6-fold; the integrity of fragments was decreased in GC compared with controls. WES analysis of cancerous tissue and plasma cfDNA (targeted sequencing) mutational profiles revealed 47.8% concordance. The increased quantity of GC tissue–derived alterations detected in cfDNA was associated with worse patients' survival. Analysis of importance of multilayer variables and receiver operating characteristic curve showed that combination of 2 analytes: (i) quantity of tissue matching alterations and (ii) presence of any somatic alteration in plasma cfDNA resulted in area under curve 0.744 when discriminating patients with or without distant metastasis. Furthermore, cfDNA sequence alterations derived from tumor tissue were detected in patients who had even relatively small GC tumors (T1-T2).**

DISCUSSION: **Our results indicate that quantitative and qualitative cfDNA mutational profile analysis is a promising tool for evaluating GC disease status or poorer prognosis.**

INTRODUCTION

Gastric cancer (GC) is one of the most common and lethal oncological diseases of the gastrointestinal tract worldwide because it is usually diagnosed at an advanced stage because of asymptomatic course of the disease (1). It is a complex disease arising from the interaction of environmental and host-associated factors (2,3), and conventional diagnostic techniques or current molecular biomarkers have a very limited role for early diagnosis of GC (4,5). Thus, minimally invasive biomarkers that would help to determine specific molecular spectra for diagnostic and prognostic purposes are highly needed.

Improving technologies have enabled a more comprehensive molecular analysis in the body fluids of patients with cancer and have revealed that circulating tumor–derived molecules could provide multilayer molecular information suitable for cancer diagnostics, prognosis, or even response to therapy (6–8). The currently available studies analyzing ctDNA alterations in GC focus on a limited number of well-known oncogenes such as *TP53* (6) and *HER2* (9–11). On the other hand, studies implementing high-throughput technologies such as new generation sequencing (NGS) are still very scarce and have been mostly conducted in Asian populations (12–14).

In this study by using cancer tissue whole exome sequencing (WES), we developed custom 38-gene panel and performed cfDNA

[1]Institute for Digestive Research, Lithuanian University of Health Sciences, Kaunas, Lithuania; [2]Institute of Clinical Molecular Biology, Christian-Albrechts-University of Kiel, Kiel, Germany; [3]Department of Gastroenterology, Lithuanian University of Health Sciences, Kaunas, Lithuania. **Correspondence:** Jurgita Skieceviciene, PhD. E-mail: jurgita.skieceviciene@lsmuni.lt.

American College of Gastroenterology

Clinical and Translational Gastroenterology

deep targeted sequencing in plasma samples. We were able to identify somatic alterations in cfDNA in a solid proportion of the patients with GC, including patients with early disease stages. Moreover, we performed multicomponent analysis for GC using machine learning on various analytes including cfDNA and oncoproteins. Our study suggests that qualitative and quantitative analysis of somatic variants in the plasma cfDNA might be a promising approach when discriminating patients based on disease state and even predict survival.

MATERIALS AND METHODS

Patient samples

Treatment-naive GC patients (n = 29) were recruited at the Department of Gastroenterology, Lithuanian University of Health Sciences Hospital during the period of 2015–2018. Clinical and demographic characteristics of patients are summarized in Figure 1 (see also Supplementary Table 1, Supplemental Digital Content 1, http://links.lww.com/CTG/A679). Paired tissue and plasma samples were collected at the same time point. Tumor tissue samples were obtained from the primary lesion during gastroscopy or surgical tumor removal. Peripheral blood was collected using K$_2$EDTA tubes (10 mL; Becton, Dickinson and Company, Franklin Lakes, NJ) for cfDNA extraction (double centrifugation protocol within 2 hours of blood draw) and serum separator tubes (5 mL; Becton, Dickinson, and Company) for serum separation. The control group (n = 20) consisted of self-reported healthy subjects without a history of cancer. All subjects provided written informed consent. Research was approved by the Kaunas Regional Biomedical Research Ethics Committee (No. BE-2-10, May 8, 2011, and No. BE-2-31, June 5, 2018, Kaunas, Lithuania).

Isolation of nucleic acids

Genomic DNA (gDNA) from the primary GC lesion was isolated using the AllPrep DNA/RNA Mini Kit (Qiagen, Hilden, Germany), and gDNA from white blood cells (WBC) was isolated using salting-out method. Total circulating nucleic acids from plasma were extracted using QIAamp Circulating Nucleic Acid isolation kit (Qiagen). All isolations of nucleic acids were performed according to the manufacturers' protocols. cfDNA yield and fragment size were evaluated using TapeStation 2200 system (Agilent Technologies, Santa Clara, CA). Tumor cfDNA fraction was calculated according to mean mutant allele frequency (MAF) in each patient's plasma sample.

Library preparation of whole exome and targeted NGS

Pair-end (2× 100 bp) sequencing libraries of GC tissue and matched WBC samples were constructed using the Illumina TruSeq Nano DNA Library Prep Kit (Illumina, San Diego, CA) according to the manufacturer's recommendations. gDNA libraries for exome sequencing were captured using the Integrated DNA Technologies xGen Exome Research Panel and hybridization reagents (Integrated DNA Technologies, Coralville, IA). Pair-end (2× 150 bp) sequencing libraries from plasma cfDNA samples were constructed using TruSight Oncology Unique Molecular Identifier (UMI) Reagents (Illumina). cfDNA was captured using Integrated DNA Technologies xGen Custom Panel consisting of 38 GC-associated mutated genes (see Supplementary Table 2, Supplemental Digital Content 2, http://links.lww.com/CTG/A680). All libraries were sequenced on the NovaSeq 6000 platform (Illumina) according to the manufacturer's instructions. The on-target sequence depth metrics are presented in Supplemental Table 3 (see Supplementary Digital Content 3, http://links.lww.com/CTG/A681).

Variant calling and development of custom gene panel for targeted sequencing

The GATK Best Practices paired-sample workflow (15) for somatic short variant discovery was used for the GC tissue exome analysis (human genome reference build hg19). Variants were called using GATK4 Mutect2 and annotated using Ensembl-VeP (v96.0) (16). Microsatellite instability (MSI) from WES data was evaluated using MSIsensor (17). Tumor mutational burden (TMB) was defined as the quantity of somatic mutations in the coding region per megabase (Mb) (18).

Filtering of somatic variants and selection of GC-related genes for cfDNA custom targeted sequencing panel was performed using following criteria: (i) prevalence of the mutation in general population <1%; (ii) protein coding nonsynonymous, annotated as having high impact; (iii) Combined Annotation Dependent Depletion score >30; (iv) excluding variants that are present in 100% of the samples; and (v) variant supported with coverage ≥2 in both forward and reverse directions.

Plasma cfDNA targeted sequencing analysis was performed using the Illumina UMI Error Correction App (v1.0.0.1), and variants were called using GATK4 Mutect2 and annotated using Ensembl-VeP (v96.0) or SnpEff (v4.3.1t) (19). All detected somatic variants were validated using the integrative genomics viewer (v2.5.3) (20).

Assessment of serum Carcinoembryonic Antigen, CA 19-9, and Cancer Antigen 72-4 level

Serum level of oncoproteins was measured by enzyme-linked immunosorbent assay (ELISA): Human Carcinoembryonic Antigen (CEA) ELISA Kit (ab99992; Abcam, Cambridge, UK), Human Cancer Antigen CA 19-9 ELISA Kit (ab108642; Abcam), and Human Cancer Antigen 72-4 (Tumor Marker CA724) ELISA Kit (E-EL-H0613; Elabscience, Wuhan, China). All analytical procedures were performed according to manufacturers' instructions.

Statistical analysis

Statistical analysis and data visualization was performed using R Studio (R version 3.3.3). Comparison of total cfDNA yield was evaluated by 2-sided *t* test or Mann-Whitney *U* test depending on the data distribution. Correlation analysis was performed using the Spearman rank-order correlation analysis. Multivariate comparison was performed using ANOVA, and 2 groups were compared using χ^2 or Fisher exact tests (2-sided). MAF analysis was conducted using maftools package (Bioconductor) (21). Gene list pathway enrichment analysis was performed using the PANTHER Gene List Analysis tool (22). Random forest analysis of the prediction variables' importance was performed using the Boruta and randomForest packages (23,24). Survival analysis was performed using the Kaplan-Meier method and Cox proportional hazards model.

RESULTS

Total cfDNA yield and size of the fragments differ between GC cases and controls

Total cfDNA yield (fragments from 100 to 1,000 bp, Figure 2a) was compared with GC clinical features and patients' characteristics. As expected, a significantly higher amount of total cfDNA was detected in patients with GC (87.59 ng per ml of plasma) compared with controls (2.01 ng per ml of plasma) (W = 0, P = 7.07×10^{14}) (Figure 2b). Moreover, the analysis of total cfDNA yield revealed positive significant correlation with serum CEA
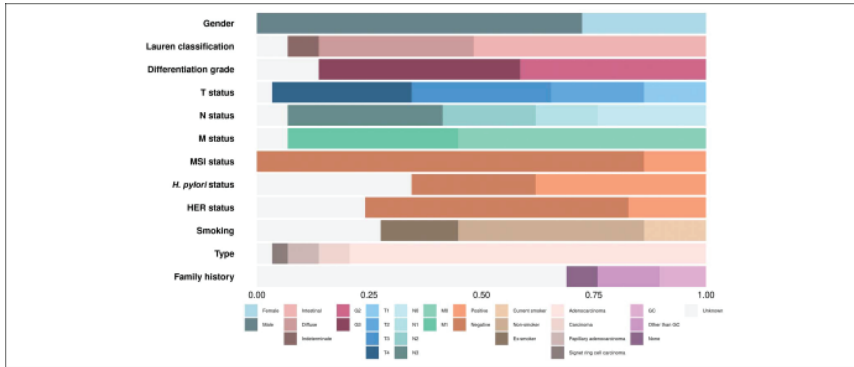
Figure 1. Characteristics of patients with GC (n = 29). A bar graph representing the proportion of patients in each section. GC, gastric cancer; HER, human epidermal growth factor receptor; MSI, microsatellite instability.

levels (see Supplementary Figure 1, Supplemental Digital Content 4, http://links.lww.com/CTG/A682).

Analysis of total cfDNA fragmentation revealed that the yield of all nucleosomal fragments was increased in the GC group compared with control: (i) mononucleosomal fragments: 61,572.60 vs 1,193.94 pg/mL (W = 0, $P = 7.07 \times 10^{-14}$); (ii) dinucleosomal fragments: 21,373.82 vs 437.80 pg/mL (W = 1, $P = 1.42 \times 10^{-13}$); and (iii) trinucleosomal fragments: 16,086.52 vs 367.03 pg/mL (W = 5, $P = 1.34 \times 10^{-12}$) (Figure 3a). The length of the fragments was shorter in the GC group compared with control: (i) mononucleosomal: 73 vs 125 bp (W = 568, $P = 1.60 \times 10^{-8}$) and (ii) dinucleosomal: 349 vs 259 bp (W = 432.5, $P = 1.06 \times 10^{-7}$) (Figure 3b).

Custom gene panel developed according to the mutational spectra of GC tissue

The GC tissue mutational spectra from WES data are presented in Supplementary Figure 2 (see Supplemental Digital Content 5, http://links.lww.com/CTG/A683). In total, 23 of 29 patients with GC (79.31%) had somatic mutations that passed the previously described selection criteria. On average, alteration-positive tissue samples had 8.4 somatic mutations (range from 1 to 23) in genes included in our custom panel. Variant allele frequencies (VAFs) ranged from 2.8% to 87.1%. Distribution of variant classifications and types is presented in Figure 4a,b. The top 10 most frequently mutated genes are shown in Figure 4f. Based on WES results, a 38-gene panel for very deep targeted sequencing of plasma cfDNA was designed. All somatic mutations of 38 genes included in our panel in tissue samples are presented in Supplementary Table 4 (see Supplemental Digital Content 6, http://links.lww.com/CTG/A684).

Mutational spectra of plasma cfDNA are associated with tumor size and survival of the patients with GC

Deep sequencing (40,000× raw coverage) of our custom gene panel was performed for plasma cfDNA samples only. Venn

diagram shows the number of detected variants in tissue and plasma (Figure 5).

Overall, somatic cfDNA alterations were observed in 21 of 23 patients with alteration-positive GC tissue samples (91.3%) (Figure 5; Supplementary Table 5 [see Supplemental Digital Content 7, http://links.lww.com/CTG/A685]). On average, 5.4 somatic variants per sample were detected in plasma cfDNA of patients with GC (range from 1 to 14). Tissue matching cfDNA alterations were detected in 11 of 23 alteration-positive GC tissue samples (47.8%) (Figure 5). On average, 3.5 tissue matching somatic variants in plasma cfDNA were detected per sample (range from 1 to 12).

Next, we compared the quantity of tissue matching somatic variants in plasma cfDNA with GC clinical features and analyzed correlation with total cfDNA yield, serum level of oncoproteins, and age. Concordantly with literature (25,26), the quantity of unique somatic alterations detected in tissue and plasma and the quantity of tissue matching alterations in plasma revealed positive moderate correlation with age (tissue: R = 0.47, P = 0.012; plasma: R = 0.4, P = 0.035; matching variants: R = 0.38, P = 0.048 [see Supplementary Figure 4, Supplemental Digital Content 8, http://links.lww.com/CTG/A686]). Our analysis revealed that cfDNA sequence alterations derived from tumor tissue were detected significantly more often in samples of the patients with larger tumors (T3-T4—55.6% and T1-T2—10.0%, $\chi^2 = 5.59$, $P = 0.018$) (Figure 6a) and in patients with distal metastasis (not significantly) (45.5% and 37.5%, M1 and M0, respectively, $\chi^2 = 0.17$, P value = 0.679) (Figure 6b). Survival analysis showed that patients without sequence alterations in cfDNA had a median survival time (MST) of 803 days, whereas MST for patients with 1–2 cfDNA sequence alterations was 469 days. MST for patients with 3–6 cfDNA sequence alterations and more than 6 cfDNA sequence alterations was 315 and 44 days, respectively (P value = 0.008) (Figure 6c). In addition, Cox proportional hazards model for the survival analysis was used. Model included not only tissue matching somatic variants detected in plasma but also patients' demographics and tumors characteristics: age, gender, and size of
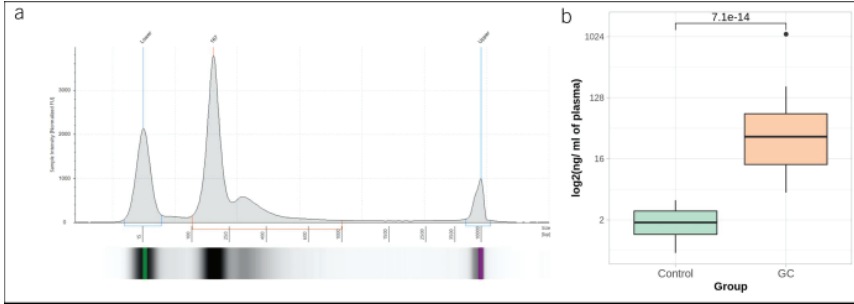
STOMACH

Figure 2. (**a**) Representative electropherogram of cfDNA sample. Fragment size of cfDNA ranges from 100 to 1,000 bp, with the main peak around 170 bp; (**b**) yield of plasma total cfDNA (ng per ml of blood plasma) in control (n = 20) and GC (n = 29) samples. A 43.6-fold increase of cfDNA yield was determined in the GC patient group ($P = 7.07 \times 10^{-14}$). Results are shown on logarithmic scale. cfDNA, cell-free DNA; GC, gastric cancer.

the primary tumor based on tumor–node–metastasis staging. Results showed slight gender impact on survival estimation (padj = 0.0410) and significant effect of more than 6 variants detected in plasma (padj = 0.0186) for shorter lifespan.

Qualitative and quantitative analysis of somatic variants in plasma discriminates patients with distant metastases
The role of multilayer molecular profiling in the discrimination of patients with larger tumors (T3-T4) and distant metastases was evaluated by including analytes such as concentration of oncoproteins CA 19-9, Cancer Antigen 72-4, CEA, MSI status, TMB, quantity of somatic mutations (unique or matching the tumor tissue), presence or absence of somatic mutations (unique or matching the tumor tissue), and specific mutations of the most mutated genes. Our analysis revealed that the quantity of tissue matching variants and the presence of any somatic alteration in plasma cfDNA was shown to be significant for discrimination

between M0 and M1 groups (classification analysis resulted in area under curve = 0.744).

DISCUSSION
In this study, we present a robust analysis of liquid biopsy for GC using circulating plasma cfDNA. We show that somatic mutations determined by WES in GC tissues can be tracked in the blood of patients with GC. Furthermore, our study suggests that qualitative and quantitative analysis of somatic variants in the plasma cfDNA might be a promising approach to discriminate patients with advanced disease.

Raised cfDNA levels were first reported in the serum of patients with cancer in 1977 (27). However, it was shown that concentration of cfDNA could increase because of number of physiological conditions, and more specific analysis of circulating nucleic acids is needed. Circulating tumor DNA can be detected in any body fluids, does not require additional analysis tools such
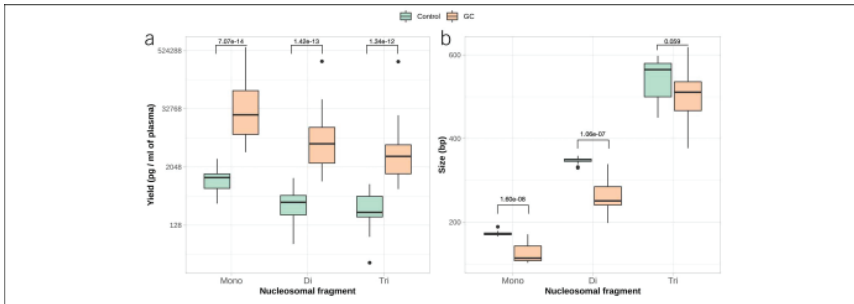

Figure 3. (**a**) Yield of mononucleosomal, dinucleosomal, and trinucleosomal fragments (pg per ml of plasma) in control and GC patients' groups. Statistically significant increase of all cfDNA peaks was observed for patients with GC ($P = 7.07 \times 10^{-14}$, $P = 1.42 \times 10^{-13}$, and $P = 1.34 \times 10^{-12}$, mononucleosomal, dinucleosomal, and trinucleosomal peaks, respectively); (**b**) size of mononucleosomal, dinucleosomal, and trinucleosomal fragments of cfDNA in control and GC patients' groups. Mononucleosomal and dinucleosomal cfDNA peaks of the patients with GC were significantly shorter compared with control cfDNA samples ($P = 1.60 \times 10^{-8}$ and $P = 1.06 \times 10^{-7}$, respectively). cfDNA, cell-free DNA; GC, gastric cancer.
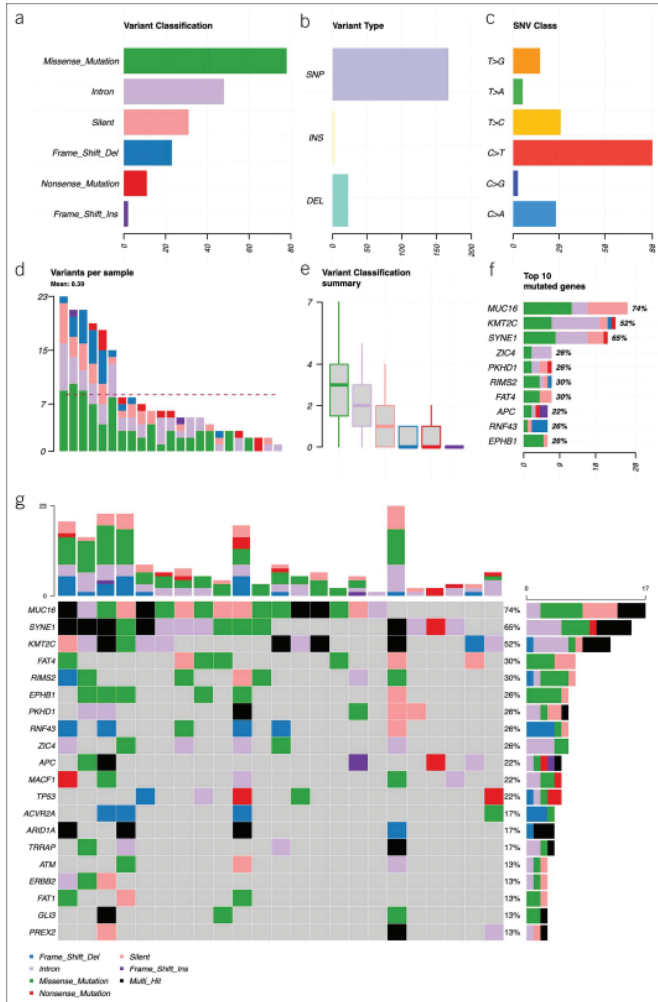
Figure 4. Summary of the mutational spectra (only genes from custom gene panel) in gastric cancer tissue samples: (**a**) absolute variant class values, the most common variant class was missense mutations; (**b**) absolute variant type values, the most common variant type detected was SNPs; (**c**) distribution of various SNV substitutions, C > T substitutions were detected the most frequently; (**d**) absolute numbers of variants per sample, the dashed line shows the mean quantity of somatic variants per sample (8.39); (**e**) mean distribution of variant classes per sample, on average, missense mutations were most frequent; (**f**) top 10 mutated genes, x axis: absolute numbers (samples), percentages calculated from all somatic variants detected; and (**g**) oncoplot of the mutated genes in gastric cancer tissue samples, showing mutated genes and distribution of variant classes per sample. Color codes in (**d-g**) graphs are the same as in (**a**). DEL, deletion; INS, insertion; SNP, single nucleotide polymorphism; SNV, single nucleotide variant.
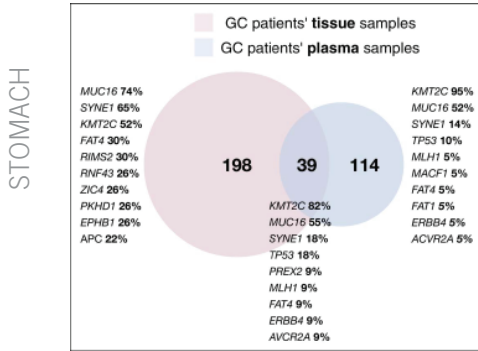
STOMACH

Figure 5. Venn diagram shows the quantity of unique and shared somatic alterations detected in GC patient tissue and plasma samples. Genes listed represents top 10 mutated genes in each case (tissue only, plasma only, and shared) and their frequency. GC, gastric cancer.

as cell sorting as in the case of circulating tumor cell analysis, and has a very high clinical potential: applications from noninvasive genomic analysis of cancer, quantification of disease burden, disease burden monitoring, and clonal evolution. Despite the recent effort (The Cancer Genome Atlas Research Network) (28), there is still high need for more appropriate gene panels for cfDNA analysis which could be implemented in the routine diagnostics. To analyze wide molecular spectra and investigate genetic alterations in the GC patient group of the European descent, we performed WES for tumor tissue and WBC samples. Twenty-three of 29 patients with GC (79.31%) had cancer-associated somatic alterations detected in tissue. All mutated genes were previously associated with gastric tumorigenesis and reported in the Catalogue Of Somatic Mutations In Cancer database (29). Signaling pathway enrichment analysis revealed that genes which we found to be mutated were involved in Wnt and cadherin pathways (see Supplementary Figure 3, Supplemental Digital Content 9, http://links.lww.com/CTG/A687) (30). Based on our WES results, a custom 38-gene panel for deep targeted sequencing of plasma cfDNA was designed. To the best of our knowledge, this is the first study conducted in patients with GC which implemented UMI error correction and deep sequencing for accurate cfDNA mutational analysis. This approach allowed us to determine somatic alterations in plasma cfDNA samples for 21 of 23 alteration-positive tissue samples (91.3%) and tumor tissue matching alterations for 11 of 23 alteration-positive tissue samples (47.8%). By comparison, previously reported plasma ctDNA mutational concordance with tissue ranged from 33.9% to 58% (8,13,14), and the differences could be explained by GC tissue molecular heterogeneity (31).

Furthermore, we have compared the quantity of tissue matching alterations detected in plasma cfDNA with different clinical features. In concordance to other studies, the analysis has revealed that alterations derived from tumor tissue were detected significantly more often in samples from the patients with more advanced tumors (6,8) and could be associated with worse survival (8). But, our data also indicated that even relatively small GC

tumors (10% of T1-T2) could shed detectable amounts of ctDNA into the blood stream. Multicomponent analysis of variable importance based on machine learning algorithms showed that combination of quantity of tissue matching alterations in cfDNA and presence of any somatic alteration in plasma cfDNA was the most accurate when discriminating patients with distal metastasis (area under curve = 0.744). However, it is important to note that more than a third of gastric tumors without distant site metastasis still gave rise to detectable cfDNA molecules carrying somatic alterations. Therefore, we believe that an ability of our custom cfDNA panel to detect even a fraction of patients with non-advanced tumors (early stages or without metastasis) could improve early cancer detection and increase survival rates (32). Studies report strong correlation between tumor-derived cfDNA detection rates and stage of tumors and in concordant with our findings show that detection rate is around 30% for tumors without distant metastasis (6,8,33,34). Moreover, survival analysis revealed that an increased quantity of somatic mutations in plasma cfDNA is associated with the worse patient's survival. Well-known cancer diagnostic analytes (MSI status, TMB, and oncoproteins) did not reveal any significant impact in our variable importance analysis or our discrimination analysis. These findings support the great need of new minimally invasive molecular markers for GC diagnosis and disease state monitoring.

In addition, we observed that the total cfDNA yield is increased in patients with GC. The higher total cfDNA yield in GC is consistent with previous studies of gastrointestinal cancers (35–38). Although results of various studies show that levels of oncoproteins such as CEA hardly correlate with clinicopathological features (39,40), we found moderate positive correlation with total cfDNA yield and serum CEA levels for patients with GC. The logical explanation for this correlation could be that increased levels of both total cfDNA yield and CEA are observed during tumorigenesis. In the analysis of cfDNA fragment distribution, we showed that the higher total cfDNA yield in GC affected all fragment sizes and mononucleosomal and dinucleosomal fragments were smaller in the patients with GC compared with the control. This observation supports the hypoxia theory: Rapidly growing tumor cells lack oxygen; hypoxia induces necrosis which leads to phagocytosis of tumor cells and DNA fragment release to the blood stream.

The study has some limitations. Study sample size is small; however, study population was well clinically defined and tested for many clinically relevant variables. Healthy controls' plasma cfDNA was not sequenced while healthy controls usually have a very low total cfDNA yield and extremely low ctDNA fraction. This could result to inconsistencies and sequencing errors. Although the gene panel was not evaluated in the independent validation group, all variants were manually checked on integrative genomics viewer. Nevertheless, we believe that this study adds very important new data for the development of clinically relevant liquid biopsy tools in patients with GC.

In conclusion, sequencing-based approaches have the advantage of being flexible and capable of detecting a wide range of aberrations in tumor genomes. Therefore, in this study, WES was performed to analyze the GC tissue mutational profile and to develop a custom panel for cfDNA mutational profile analysis. It is important to note that by using our gene panel and UMI correction, we were able to detect tumor-derived cfDNA even for small tumors and tumors without distant metastasis and identify a solid proportion of patients with GC carrying somatic
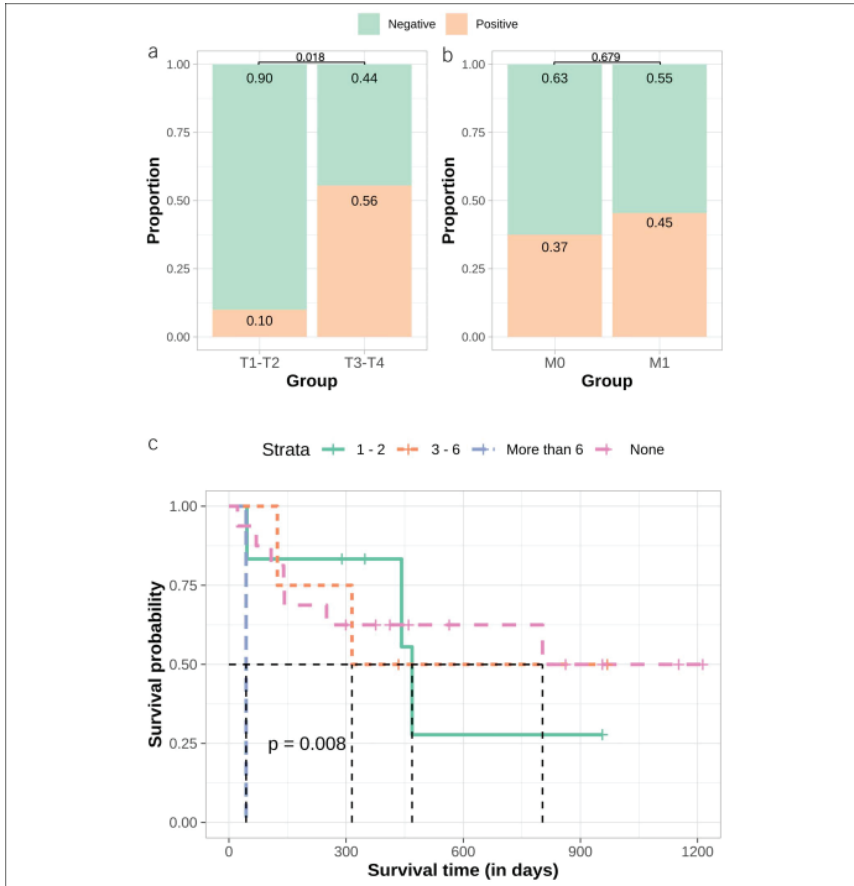
STOMACH

Figure 6. (**a**) Proportion of positive and negative samples with matching tissue and plasma cfDNA alterations comparing groups of T1-T2 and T3-T4. Approximately 10.0% of T1-T2 GC patients and 56.0% of T3-T4 GC patients had a detectable amount of circulating tumor DNA; difference was statistically significant ($P = 0.018$); (**b**) proportion of positive and negative samples with matching tissue and plasma cfDNA alterations comparing groups of M0 and M1. Approximately 37.0% of the GC without distant metastasis and 45.0% of GC with distant metastasis had a detectable amount of tumor cfDNA; difference was not significant ($P = 0.679$); (**c**) Kaplan-Meier survival analysis of patients with GC with 0 (pink line), 1–2 (green line), 3–6 (orange line), or more than 6 (blue line) tissue matching alterations detected in plasma cfDNA. Average survival in days decreases gradually when comparing patients with increasing quantity of mutations ($P = 0.008$). cfDNA, cell-free DNA; GC, gastric cancer.

alterations in plasma cfDNA. We found that the quantity of somatic alterations could be associated with overall patients' survival. Further investigation of plasma cfDNA could implement larger cohorts of the patients with GC and analysis of MAF in cfDNA at different disease time points and/or disease status (e. g.

relapse or remission). The implementation of plasma cfDNA analysis into routine cancer testing is still technically challenging, and more population-based screening studies are still needed. However, given the progress in NGS technology and new methods of processing complex data, tumor-derived cfDNA even

today shows potential clinical utility as a noninvasive analyte for the characterization of an individual patient's tumor genome.

Study Highlights

WHAT IS KNOWN

✓ gastric cancer (GC) diagnosis in late stages and high mortality rates indicate the need for new molecular tools.
✓ Conventional diagnostic techniques or current molecular biomarkers have a very limited role for early diagnosis of GC.
✓ Tumor-derived DNA (ctDNA) found in plasma of patients with cancer carry genetic information of the tumor which could be assessed by minimally invasive way.

WHAT IS NEW HERE

✓ By using GC tissue whole exome sequencing, we developed a custom 38-gene panel and performed cfDNA deep targeted sequencing in plasma samples.
✓ This unique gene panel enabled to identify a solid proportion of patients with GC carrying somatic alterations in plasma cfDNA, including patients whose disease was in early stages.
✓ Multilayer molecular machine learning–based analysis indicated that the quantity of tissue matching variants and the presence of any somatic alteration in plasma cfDNA is significant for discrimination between M0 and M1 groups.

REFERENCES

1. Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018;68:394–424.
2. Rajilic-Stojanovic M, Figueiredo C, Smet A, et al. Systematic review: Gastric microbiota in health and disease. Aliment Pharmacol Ther 2020;51:582–602.
3. Zhao Y, Zhang J, Cheng ASL, et al. Gastric cancer: Genome damaged by bugs. Oncogene 2020;39:3427–42.
4. Bornschein J, Leja M, Kupcinskas J, et al. Molecular diagnostics in gastric cancer. Front Biosci 2014;19:312–38.
5. Link A, Kupcinskas J. MicroRNAs as non-invasive diagnostic biomarkers for gastric cancer: Current insights and future perspectives. World J Gastroenterol 2018;24:3313–29.
6. Hamakawa T, Kukita Y, Kurokawa Y, et al. Monitoring gastric cancer progression with circulating tumour DNA. Br J Cancer 2015;112: 352–6.
7. Kim K, Shin DG, Park MK, et al. Circulating cell-free DNA as a promising biomarker in patients with gastric cancer: Diagnostic validity and significant reduction of cfDNA after surgical resection. Ann Surg Treat Res 2014;86:136–42.
8. Fang WL, Lan YT, Huang KH, et al. Clinical significance of circulating plasma DNA in gastric cancer. Int J Cancer 2016;138:2974–83.
9. Shoda K, Ichikawa D, Fujita Y, et al. Monitoring the HER2 copy number status in circulating tumor DNA by droplet digital PCR in patients with gastric cancer. Gastric Cancer 2017;20:126–35.
10. Aguilar-Mahecha A, Joseph S, Cavallone L, et al. Precision medicine tools to guide therapy and monitor response to treatment in a HER-2+ gastric cancer patient: Case report. Front Oncol 2019;9:698.
11. Kinugasa H, Nouso K, Tanaka T, et al. Droplet digital PCR measurement of HER2 in patients with gastric cancer. Br J Cancer 2015; 112:1652–5.
12. Shu Y, Wu X, Tong X, et al. Circulating tumor DNA mutation profiling by targeted next generation sequencing provides guidance for personalized treatments in multiple cancer types. Sci Rep 2017;7:583–11.
13. Kim YW, Kim YH, Song Y, et al. Monitoring circulating tumor DNA by analyzing personalized cancer-specific rearrangements to detect recurrence in gastric cancer. Exp Mol Med 2019;51:1–10.
14. Lan J, Lu Y, Guan Y, et al. Identification of circulating tumor DNA using a targeted 545-gene next generation sequencing panel in patients with gastric cancer. Oncol Lett 2020;19:2251–7.
15. McKenna A, Hanna M, Banks E, et al. The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010;20:1297–303.
16. McLaren W, Gil L, Hunt SE, et al. The Ensembl variant effect predictor. Genome Biol 2016;17:122–14.
17. Niu B, Ye K, Zhang Q, et al. MSIsensor: Microsatellite instability detection using paired tumor-normal sequence data. Bioinformatics 2014;30:1015–6.
18. Xu Z, Dai J, Wang D, et al. Assessment of tumor mutation burden calculation from gene panel sequencing data. Onco Targets Ther 2019;12: 3401–9.
19. Cingolani P, Platts A, Wang LL, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of drosophila melanogaster strain w1118; iso-2; iso-3. Fly 2012;6:80–92.
20. Robinson JT, Thorvaldsdóttir H, Winckler W, et al. Integrative genomics viewer. Nat Biotechnol 2011;29:24–6.
21. Mayakonda A, Lin DC, Assenov Y, et al. Maftools: Efficient and comprehensive analysis of somatic variants in cancer. Genome Res 2018; 28:1747–56.
22. Mi H, Muruganujan A, Ebert D, et al. PANTHER version 14: More genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. Nucleic Acids Res 2019;47:D419–D426.
23. Kursa MB, Rudnicki WR. Feature selection with the boruta package. J Stat Softw 2010;36:1–13.
24. Liaw A, Wiener M. Classification and regression by RandomForest. R News 2002;2:18–22.
25. Milholland B, Auton A, Suh Y, et al. Age-related somatic mutations in the cancer genome. Oncotarget 2015;6:24627–35.
26. Grist SA, McCarron M, Kutlaca A, et al. In vivo human somatic mutation: Frequency and spectrum with age. Mutat Res 1992;266: 189–96.
27. Leon SA, Shapiro B, Sklaroff DM, et al. Free DNA in the serum of cancer patients and the effect of therapy. Cancer Res 1977;37:646–50.
28. Bass AJ, Thorsson V, Shmulevich I, et al. Comprehensive molecular characterization of gastric adenocarcinoma. Nature 2014;513:202–9.
29. Tate JG, Bamford S, Jubb HC, et al. COSMIC: The catalogue of somatic mutations in cancer. Nucleic Acids Res 2019;47:D941–D7.
30. Nelson WJ, Nusse R. Convergence of Wnt, β-catenin, and cadherin pathways. Science 2004;303:1483–7.
31. Stahl P, Seeschaaf C, Lebok P, et al. Heterogeneity of amplification of HER2, EGFR, CCND1 and MYC in gastric cancer. BMC Gastroenterol 2015;15:7–13.

32. Cho H, Mariotto AB, Schwartz LM, et al. When do changes in cancer survival mean progress? The insight from population incidence and mortality. J Natl Cancer Inst 2014;49:187–97.
33. Bettegowda C, Sausen M, Leary RJ, et al. Detection of circulating tumor DNA in early- and late-stage human malignancies. Sci Transl Med 2014;6:224ra24.
34. Cabel L, Decraene C, Bieche I, et al. Limited sensitivity of circulating tumor DNA detection by droplet digital PCR in non-metastatic operable gastric cancer patients. Cancers 2019;11:396.
35. Sai S, Ichikawa D, Tomita H, et al. Quantification of plasma cell-free DNA in patients with gastric cancer. Anticancer Res 2007;27:2747–51.
36. Park JL, Kim HJ, Choi BY, et al. Quantitative analysis of cell-free DNA in the plasma of gastric cancer patients. Oncol Lett 2012;3:921–6.
37. Hsieh CC, Hsu HS, Chang SC, et al. Circulating cell-free DNA levels could predict oncological outcomes of patients undergoing esophagectomy for esophageal squamous cell carcinoma. Int J Mol Sci 2016;17:2131.
38. Mouliere F, El Messaoudi S, Pang D, et al. Multi-marker analysis of circulating cell-free DNA toward personalized medicine for colorectal cancer. Mol Oncol 2014;8:927–41.
39. Lai IR, Lee WJ, Huang MT, et al. Comparison of serum CA72-4, CEA, TPA, CA19-9 and CA125 levels in gastric cancer patients and correlation with recurrence. Hepatogastroenterology 2002;49:1157–60.
40. Qian C, Ju S, Qi J, et al. Alu-based cell-free DNA: A novel biomarker for screening of gastric cancer. Oncotarget 2017;8:54037–45.

CURRICULUM VITAE

Name, Surname:	**Greta Varkalaitė**
Mobile:	+370 643 12903
E-mail address:	greta.varkalaite@lsmuni.lt

Education:

2017–2021 Lithuanian University of Health Sciences, Kaunas, Lithuania
Ph. D. student in Biomedical Sciences

2015–2017 Lithuanian University of Health Sciences, Kaunas, Lithuania
M.S. in Biology of Laboratory Medicine

2011–2015 Lithuanian University of Health Sciences, Kaunas, Lithuania
B.S. in Medical and Veterinary Genetics

Research experience:

2017–present Institute for Digestive Research, Laboratory of Clinical and
Molecular Gastroenterology, Lithuanian university of health sciences,
Kaunas, Lithuania
Junior Scientist

2016–2017 Institute for Digestive Research, Laboratory of Clinical and
Molecular Gastroenterology, Lithuanian university of health sciences,
Kaunas, Lithuania
Laboratory technician

Research projects:

12/2021–present Project "SARS-CoV-2 genomics: providing a tool for viral evolution
rate assessment, to inform exploratory modelling of immune and
clinical response (evoCOVID)" (No. 01.2.2-LMT-K-718-05-0026).
Project leader: prof. Limas Kupčinskas, duration: 2021-2023; LSMU.
Junior Scientist

05/2021–present Project "Gut-blood-liver axis: Circulating microbiome as non-
invasive biomarker for Inflammatory bowel disease (IBD) and
Primary sclerosing cholangitis (PSC) (Blood-Biome)"
(No. S-BMT-21-11 (LT08-2-LMT-K-01-060)). Project leader:
Prof. Habil. Dr. Gediminas Kiudelis, duration: 2021 – 2023; LSMU.
Junior Scientist

06/2020–present Project "Intestinal Organoid Model: Platform for Drug Testing and
Personalized Therapy in Inflammatory bowel diseases
(IBD.ORG)" (No. LMT-K-718-04-0003). Project leader:
prof. Jurgita Skiecevičienė, duration: 2020 – 2023; LSMU.
Junior Scientist

01/2018–present	Project "The multi-stage molecular study of gastric cancer: blood circulating genetic and epigenetic biomarkers and profiling of microbiome biomarkers (MULTIOMICS)" (No. 09.3.3-LMT-K-712-01-0130). Project leader: prof. Jurgita Skiecevičienė, duration: 2018-2021; LSMU. ***Junior Scientist***
06/2016–06/2019	Project "Gastric microbiome: twin study" (Nr. APP-2/2016). Project leader: prof. Jurgita Skiecevičienė, duration: 2016-2019; LSMU. ***Laboratory Technician***
08/2015–12/2017	Project "Healthy and patients with age-related macular degeneration new genetic and metabolic markers search" (Nr. SEN-11/2015). Project leader: prof. V. Deltuva, duration: 2015-2018; LSMU. ***Laboratory Technician***
03/2017–06/2017	Project "Volatile biomarkers for early detection and characterization of gastric and colorectal Neoplasms (VOLGACORE)" (Nr. EURONANOMED2-01/2014). Project leader: prof. Limas Kupčinskas, duration: 2014-2017; LSMU. ***Superior Laboratory Technician***
04/2015–12/2016	Project "MicroRNA profiling and functional analysis in gastrointestinal stromal tumors" (Nr. MIP-006/2014). Project leader: prof. Jurgita Skiecevičienė, duration: 2014-2016; LSMU. ***Laboratory Technician***

International experience:

10/2017-11/2017	Christian-Albrechts University of Kiel, Institute of Clinical Molecular Biology, Kiel, Germany ***Scientific internship***
01/2019-03/2019	Christian-Albrechts University of Kiel, Institute of Clinical Molecular Biology, Kiel, Germany ***Scientific internship***

Scientific achievements

2022 – National Scholarship of World Federation of Scientists (WFS).

2022 – Selected to participate in the 71st Lindau Nobel Laureate Meeting (dedicated to Chemistry).

2020 – Promotional scholarship of Research Council of Lithuania for PhD students in science, actively conducting scientific research (No. P-DAP-20-14).

2019 – UEG National Scholar Award for oral presentation at UEG Week Barcelona (October 2019).

2019 – Project funded by EU Structural Funds according to the 2014–2020 Operational Programme for the European Union Funds' Investments priority "Development of Scientific Competences in scientific internship" (No. 09.3.3.-LMT-K-712-14-0098, February 2019 – March 2019).

2019 – Grant for doctoral academic visits of Research Council of Lithuania (No. P-DAK-19-152, participation in international conference UEG Week 2019).

2019 – Promotional scholarship of Research Council of Lithuania for PhD students in science, actively conducting scientific research (No. P-DAP-19-109).

2018 – Grant for doctoral academic visits of Research Council of Lithuania (No. P-DAK-18-187, participation in international conference UEG Week 2018).

2018 – Promotional scholarship of Research Council of Lithuania for PhD students in science, actively conducting scientific research (No. P-DAP-18-83).

2017 – Best oral presentation in international conference "Fit for the future in GI" (November 2017).

2016 – Nominal scholarship of ex-EU commissioner and WHO's special envoy for Europe Mr. Vytenis Povilas Andriukaitis.

Supervision of students' work:

Supervisor of Additional Practice at LSMU (June-July 2021).

Master's Thesis Consultant. Medical genetics, University of Vilnius (VU), work defended in June 2021 (title: "The role of hsa-miR-196a and hsa-miR-129-1 in the pathogenesis of gastric cancer").

Master's Thesis Consultant. Laboratory medical biology, LSMU, work defended in June 2021 (title: "Cell-Free DNA Analysis in Plasma of Gastric Cancer Patients").

Supervisor of bachelor thesis. Medical and veterinary genetics, LSMU, work defended in January 2021 (title: "MicroRNA profiling of atrophic gastritis and gastric cancer tissue").

Supervisor of Practice of professional activity (Medical and veterinary genetics, LSMU), September-October 2020.

Supervisor of bachelor thesis. Medical and veterinary genetics, LSMU, thesis defended in January 2019 (title: "Expression level evaluation of miR-1468-5p and potential target genes in gastric cancer").

Supervisor of Additional Practice at LSMU (July-August 2018).

Supervisor of Practice of professional activity (Medical and veterinary genetics, LSMU), September-October 2018.

ACKNOWLEDGEMENTS

First, I would like to acknowledge and give my biggest thanks to my supervisor prof. **Jurgita Skiecevičienė** for her continued support and guidance through all stages of the preparation of this dissertation. She is not only a great leader of the group but also a huge inspiration. Second, I would like to express my deep gratitude to the former head of the institute prof. **Limas Kupčinskas** and present head of the institute prof. **Juozas Kupčinskas** for the trust, support, and opportunity to work and grow as a scientist.

Next, I would like to sincerely thank my supervisor of the scientific internships dr. **Michael Forster** for his genuine care, patience, and valuable scientific advices. Also, I would like to extend my special thanks to prof. **Andre Franke** for hosting me in the Institute of Clinical Molecular Biology (IKMB, Kiel, Germany) and providing the possibility to learn state-of-the-art methods in molecular biology and applied bioinformatics. This work could not have been completed without co-operation and most importantly huge support from our group members in Institute for Digestive Research (Kaunas, Lithuania) including dr. **Simonas Juzėnas**, dr. **Ugnė Kulokienė, Rūta Inčiūraitė, Rokas Lukoševičius, Justina Arštikytė,** and members of IKMB, including **Regina Frederik**. I especially thank dr. **Rūta Steponaitienė** for being my first mentor and for guiding me on the right path at the very beginning of this journey.

Finally, I would like to express my appreciation to my family (especially my mother and my life partner Rolandas), and dearest friends for unconditional love, support, and belief in me. All this, motivated and encouraged me to set ambitious goals, and allowed me to build the confidence in myself. I have always felt your presence in the most successful or less successful moments of my life.